PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 34 February 15, 1948 Number 2

SOME REMARKS ON THE WEIGHTS OF UNKNOWNS AS DETERMINED BY THE METHOD OF DIFFERENTIAL CORRECTIONS

By S. L. Piotrowski

Harvard College Observatory

Communicated by Harlow Shapley, December 10, 1947

In many problems of practical astronomy we apply the method of least square to non-linear equations. The computation is usually done by the method of differential corrections. If the equations of condition are

$$f_i(x, y, \ldots) = L_i, \qquad i/1 \ldots n, \tag{1}$$

where x, y ... are the unknown parameters to be determined and L_i are the observed quantities, the procedure is the following. We start with certain approximate values for x, y—let us denote them by x_0, y_0. For simplicity we limit our discussion to the case of two unknown parameters. We determine the corrections Δx and Δy by the least squares method from the equations

$$\frac{\partial f_i}{\partial x}\Delta x + \frac{\partial f_i}{\partial y}\Delta y = L_i - f_i, \tag{2}$$

with x and y in f_i, $\partial f_i/\partial x$, $\partial f_i/\partial y$ equal to x_0, y_0. We add Δx and Δy to x_0 and y_0 and, in theory at least, we repeat this procedure until the resulting corrections become equal zero. We shall assume in what follows that this can be achieved. In other words, we make the assumption that the whole procedure is convergent. It is easily seen that the values of x and y which we obtain in this way satisfy the equations

$$\left.\begin{aligned} \sum_{i/1}^{n} (L_i - f_i)\frac{\partial f_i}{\partial x} &= 0 \\ \sum_{i/1}^{n} (L_i - f_i)\frac{\partial f_i}{\partial y} &= 0. \end{aligned}\right\} \tag{3}$$

Equations (3) are, in fact, nothing but the necessary conditions making the sum of the squares of the residuals a minimum.

When solving (2) by the method of least squares we get the mean errors of the corrections Δx and Δy. These errors are characteristic for Δx and Δy only in so far as we regard Δx and Δy as defined by the normal equations pertaining to equations (2). In almost all cases known to me it is assumed (at least tacitly) that these errors represent also the mean errors of the solution obtained for x and y. It can be easily shown that, in general, this assumption is not valid.*

The inverse weights of Δx and Δy are equal to

$$\sum_{j/1}^{n}\left(\frac{\partial \Delta x}{\partial L_j}\right)^2 \quad \text{and} \quad \sum_{j/1}^{n}\left(\frac{\partial \Delta y}{\partial L_j}\right)^2,$$

respectively. We determine $\partial \Delta x/\partial L_j$ and $\partial \Delta y/\partial L_j$ by differentiating with respect to L_j the normal equations pertaining to equations (2). We get (the square brackets are used to denote the summation over i from 1 to n)

$$\left.\begin{aligned}\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial x}\right]\frac{\partial \Delta x}{\partial L_j}+\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right]\frac{\partial \Delta y}{\partial L_j}&=\frac{\partial f_j}{\partial x},\\ \left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right]\frac{\partial \Delta x}{\partial L_j}+\left[\frac{\partial f}{\partial y}\frac{\partial f}{\partial y}\right]\frac{\partial \Delta y}{\partial L_j}&=\frac{\partial f_j}{\partial y}.\end{aligned}\right\} \tag{4}$$

By a well-known theorem[1] we find the inverse weights as the diagonal elements of the inverse of the matrix of the coefficients of equations (4).

In order to find the quantities $\partial x/\partial L_j$ and $\partial y/\partial L_j$ that are prerequisite for the determination of the weights of x and y we differentiate equations (3) with respect to L_j and we get

$$\left.\begin{aligned}\left\{\left[\frac{\partial^2 f}{\partial x^2}(f-L)\right]+\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial x}\right]\right\}\frac{\partial x}{\partial L_j}+\left\{\left[\frac{\partial^2 f}{\partial x \partial y}(f-L)\right]+\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right]\right\}\frac{\partial y}{\partial L_i}&=\frac{\partial f_j}{\partial x}\\ \left\{\left[\frac{\partial^2 f}{\partial x \partial y}(f-L)\right]+\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right]\right\}\frac{\partial x}{\partial L_j}+\left\{\left[\frac{\partial^2 f}{\partial y^2}(f-L)\right]+\left[\frac{\partial f}{\partial y}\frac{\partial f}{\partial y}\right]\right\}\frac{\partial y}{\partial L_j}&=\frac{\partial f_i}{\partial y}.\end{aligned}\right\} \tag{5}$$

Now, by comparing (4) and (5) we see that the coefficients of the unknowns are not identical in both systems. They are identical when $[(\partial^2 f/\partial x^2)(f - L)]$, $[(\partial^2 f/\partial x \partial y)(f - L)]$, $[(\partial^2 f/\partial y^2)(f - L)]$ equal zero. But these conditions are not fulfilled in general. Even when the residuals $f_i - L_i$ are small and do not show any systematic behavior the considered terms

may be, in cases, quite appreciable. If we have used equations (4) for the determination of the mean errors of Δx and Δy, we should check whether the sums $[(\partial^2 f/\partial x^2)(f - L)]$, $[(\partial^2 f/\partial x \partial y)(f - L)]$, ... are really negligible before we ascribe these errors to the unknowns themselves.† That this precaution is essential, we can see from the following numerical example.

Let us suppose that for 5 equidistant values of t: 0, 1, 2, 3, 4, we have observed the values of L, which is supposed to be given by the expression

$$tx + (t - 1)^2 y - (t - 1)(t - 3)xy = L. \qquad (6)$$

Let the observed values be: -8.0, $+4.0$, $+9.5$, $+13.5$, $+11.0$. After some trials, we find by the method of differential correction that the solution is: $x = 3$, $y = 1$. The residuals are: 0.0, $+1.0$, -0.5, $+0.5$, -1.0. The matrix of the coefficients of the system (4) is

$$\begin{pmatrix} 29, & 48 \\ 48, & 96 \end{pmatrix}.$$

Its inverse is

$$\begin{pmatrix} 0.2, & -0.1 \\ -0.1, & 0.0604 \end{pmatrix}.$$

Hence, denoting by $p_{\Delta x}$ the weight of Δx, by $p_{\Delta y}$ the weight of Δy (with $x = 3$, $y = 1$, both Δx and Δy are equal 0) we obtain

$$p_{\Delta x}^{-1} = 0.2; \qquad p_{\Delta y}^{-1} = 0.0604. \qquad (7)$$

If we would ascribe these weights to x and y themselves (as is generally done), we would run into a contradiction. We can write (6) in the form

$$A + Bt + Ct^2 = L, \qquad (8)$$

with

$$A = y(1 - 3x), \qquad B = x - 2y + 4xy, \qquad C = y(1 - x). \quad (9)$$

Eliminating x and y between the three equations (9), we get

$$A - C - AB + 3CB + 2AC - A^2 + 3C^2 = 0. \qquad (10)$$

Now we can treat A, B, C, as auxiliary (unknown) parameters of the problem and determine them by the least squares method from equations (8) linear in A, B, C, ($t = 0, 1, \ldots$; $L = -8.0, +4.0, \ldots$) with the condition (10) to be satisfied exactly. It may be done by a standard method[2] and we get (returning from A, B, C, to x and y), $x = 3$, $y = 1$, the same values obtained by the method of differential corrections. The weights of x and y, p_x, p_y are, however

$$p_x^{-1} = 0.0916; \qquad p_y^{-1} = 0.0277, \qquad (11)$$

more than two times greater than those given by (7). The discrepancy is of course spurious. If we determine $\partial x/\partial L_j$, $\partial y/\partial L_j$ from equations (5) and form the sums of their squares we obtain exactly the values given by (11).

Summary.—If one uses the method of differential corrections in a least squares solution, the mean errors of the differential corrections to the unknowns are equal to the mean errors of the unknowns themselves only in the special case when the sums of the products of the residuals by the second order partial derivatives of the functions figuring in the equations of the problem are negligible. This is so regardless of how small the differential corrections happen to be. If the sums are not negligible, the equations of the form (5) should be used when determining the weights of the unknowns.

* An extensive discussion of this and related problems is to be found in a paper by E. B. Wilson and R. R. Puffer, "Least Squares and Laws of Population Growth," *Proc. Amer. Acad. Arts Sci.*, **68**, No. 9 (1933). Cf., in particular, equations (25) and (26) and the considerations in the Appendix.

† Evidently the functions of the type $f(x, y, t)$ which fulfill the system of equations

$$\partial^2 f/\partial x^2 = a(x, y)\,(\partial f/\partial x) + b(x, y)\,(\partial f/\partial y),$$
$$\partial^2 f/\partial x \partial y = a'(x, y)\,(\partial f/\partial x) + b'(x, y)\,(\partial f/\partial y),$$
$$\partial^2 f/\partial y^2 = a''(x, y)\,(\partial f/\partial x) + b''(x, y)\,(\partial f/\partial y),$$

where $a, b;\ a', b';\ a'', b''$ are arbitrary functions of x and y, independent of t, will have the property of yielding in the standard least squares solution for differential corrections the correct values of the mean errors (t is used instead of subscript i).

[1] Whittaker, E. T., and Robinson, G., *The Calculus of Observations*, London, 1932, p. 241.

[2] *Ibid.*, p. 252.

GALACTIC AND EXTRAGALACTIC STUDIES, XVIII. THE PULSE INDEX FOR EIGHTY-NINE VARIABLE STARS IN THE MAGELLANIC CLOUDS

BY HARLOW SHAPLEY

HARVARD COLLEGE OBSERVATORY

Communicated December 9, 1947

The Large and Small Clouds of Magellan, because of their isolation from the confusion of Milky Way phenomena, and their distances of only 25 kiloparsecs[1] are ideally situated for the study of the various characteristics of Cepheid variable stars. A fair sampling of the variable star population, suitable for detailed studies, can be taken from the more than 2500 variable stars in the two Clouds. It is thus possible in selecting a representative sample to choose stars with almost any required period-length and avoid those that may be photometrically complicated by companion stars or by dark or bright nebulosity.

The correlations with each other of periods, amplitudes, absolute luminosities and light-curve forms are more satisfactorily examined in the Clouds than in any other region of the sky. The advantages include the negligibility of the distance differentials for the various Cepheids, the apparent brightness of the variables compared with those in other external systems, and the possibility of using for each Cloud a single set of magnitude standards. The light curves can therefore be obtained on a homogeneous basis, both as to the zero point and the magnitude scale.

In an earlier paper of this series[2] a report was made on 40 variables of the Large Magellanic Cloud. It was shown that the amplitude, with numerous exceptions, varied with period-length; that the form of the light curve was not closely dependent on period (Hertzsprung Effect) except in the vicinity of periods of ten days; and that the period-luminosity relation appeared as usual, as did also the genuine scatter of values, about the mean period-luminosity curve.

A study of a comparable selection of Cepheid variables in the Small Cloud has been undertaken to reinforce, with somewhat richer observational material, the earlier data on the general characteristics of Cepheid light curves, and to take advantage of the still better conditions in the compact Small Cloud for homogeneous photometry.

The observations, the light curves, and a full discussion will be published elsewhere. In this communication a report is made on the asymmetry of the light curves from both Clouds. These curves are of course indicators of the nature of the underlying periodic pulsations.

The prevalent asymmetry in Cepheid light curves can be, and for the Cepheids of the galactic system have been, variously described quanti-

tatively, most commonly in terms of time proportions; for example, by the ratio

$$(t \text{ max.} - t \text{ min.})/\text{period}.$$

S. Gaposchkin[3] and others have used the ratios of areas or their differences under specific sections of the light curves as criteria of the degree of asymmetry. C. P. Gaposchkin recently has applied harmonic analysis to a number of Cepheid light curves.[4]

The time of maximum of a Cepheid of short period can be determined with fair accuracy, but the variations at minimum are so slow that even for good light curves the instant of minimum, or the first beginning of the upswing in light, cannot be accurately recorded. The middle of the rise to maximum, however, can be the most accurately observed part of the light curve. A single simple parameter that defines accurately this crucial phase of the pulsation can be obtained from the steepness of the rise; and this pulse-index, if suitably derived, is essentially independent of amplitude of variation and thus is not vitiated by unrecognized optical doubling or by absorption or background effects.

For the derivation of the pulse-index from the light curve, we avoid the slow magnitude variation at minimum, and irregularities like double or hesitant maxima, by defining the relation as follows:

$$\text{p.i.} = \frac{P}{t_{M'} - t_{m'}},$$

where the period, P, and the times, t, are expressed in the same unit (in days; or better, in period lengths, in which case the numerator is unity), and where M' is the photographic magnitude two tenths of the amplitude fainter than maximum, and m' is the photographic magnitude two tenths of the amplitude brighter than the minimum. (Relative steepness increases with p.i.) By involving only the steepest six tenths of the rise from minimum to maximum in the determination of the index, we avoid completely the uncertainties of curve drawing and the minor fluctuations at maxima and minima.

Various other indices could be used, such as the magnitude change per hour at median magnitude on the rising branch; or the magnitude change per specified fraction of the period; or the ratio of the pulse index as defined above to a comparable index for the descending branch; or smaller fractions of the amplitude might underlie the measures of the pulse index; or intensities might be substituted for magnitudes. No clear advantage, however, would come from these alternatives. In particular, the irregularities that are frequent on the descending branch of a typical Cepheid light curve argue against the use of the descending branch for a quanti-

tative index. Moreover, it can be shown, as in figure 1, that the "recovery-index" derived from the descending branch—r.i. $= P/(t_{m'} - t_{M'})$—has a small spread and in consequence is not a sensitive parameter for describing a light curve. Consequently the ratio of p.i. to r.i. has no advantages over the pulse index. In figure 1, lower array, the outstanding r.i. at 9.5 refers to a variable with double maximum, a steep decline, and a period of 9.4 days. In the upper array, which shows the wide spread of p.i., the double maxima light curves are indicated by arrows. The difference in the distribution of p.i. values for the two Clouds is probably not significant. It is most noticeable around p.i. = 6.

In table 1 the pulse index and other data are given for both Clouds, the variables arranged in order of increasing period. The Small Cloud stars are marked with asterisks.

In figure 2 the pulse index is plotted against the logarithm of the period, with dots for the Small Cloud, open circles for the Large and crosses showing the means of ten points. The computed correlation coefficient is

$$r\ (\text{p.i., log } P) = 0.44,$$

and for absolute magnitude,

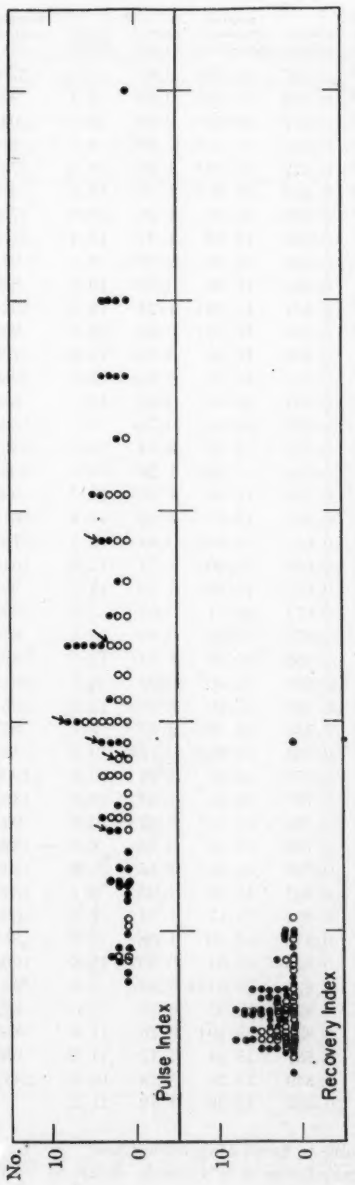

FIGURE 1

Asymmetry in light curves of 89 Cepheid Variables. Open circles, Large Magellanic Cloud; dots, Small Magellanic Cloud.

TABLE 1

HARVARD VARIABLE NO.	LOG OF PERIOD	MEDIAN APPARENT MAGNITUDE	AMPLITUDE	PULSE INDEX	HARVARD VARIABLE NO.	LOG OF PERIOD	MEDIAN APPARENT MAGNITUDE	AMPLITUDE	PULSE INDEX
1505*	0.097	16.42†	0.81	15.4	2752	0.857	16.41	1.26	8.7
1513*	0.134	16.55†	0.60	6.1	927	0.877	15.47	1.20	7.7
10364*	0.177	16.57:	0.96:	25.0:	2081*	0.881	15.10†	0.74	6.9
2127*	0.263	16.10†	0.61	8.0	816*	0.900	15.61†	0.80	7.4
1446*	0.272	16.62†	1.25	18.2	2722	0.905	15.10	1.26	8.3
2128*	0.356	16.01†	1.06	18.2	1338*	0.929	15.21†	1.24	11.8
2076*	0.398	16.12	1.24	20.0	1790*	0.948	15.28	0.88	9.5
2809	0.399	15.94	1.17	15.4	2103*	0.953	15.49†	0.54	9.5
2344	0.402	16.29	0.66	8.7	971	0.968	15.11	1.10	10.0
2796	0.461	16.26	0.93	10.0	836*	0.973	15.22†	1.05	14.3
1460*	0.464	15.98†	1.24	18.2	1334*	0.975	15.10†	1.26	10.0
2053*	0.508	16.16†	0.83	20.0	952	0.981	15.48	1.07	9.1
2368	0.508	16.36	0.89	12.5	1768*	0.992	15.33	0.43	6.2
2361	0.512	16.51	0.95	13.3	2060*	1.008	14.83	0.96	7.4
11206*	0.531	16.07†	0.68	15.4	818*	1.014	15.15†	1.26	11.8
2472	0.557	16.10	0.78	9.5	1426*	1.019	15.45	0.99	4.9
2085*	0.570	16.29	0.94	13.3	2432	1.038	14.66	0.62	9.1
858*	0.583	15.82†	1.21	20.0	2063*	1.048	15.16†	0.90	4.4
2795	0.593	16.03	1.10	16.7	999	1.051	14.86	0.59	11.8
2788	0.639	15.67	1.05	10.5	2017*	1.057	14.99†	1.12	4.0
1619*	0.641	15.98†	0.83	14.3	2787	1.058	14.99	1.03	4.3
1425*	0.658	15.90†	1.21	11.8	1610*	1.066	15.24†	1.21	3.8
851*	0.671	15.83†	1.19	18.2	905	1.074	14.96	1.40	4.4
2334	0.671	16.21	0.91	7.7	857*	1.079	14.77†	1.22	6.2
2826	0.673	15.83	0.89	15.4	856*	1.085	15.28†	1.12	16.7
951	0.688	16.16	1.01	12.5	1365*	1.094	15.18†	1.11	6.1
2861	0.694	16.00	0.99	8.7	2052*	1.099	14.55†	1.12	4.3
2031*	0.720	15.60	1.29	12.5	1351*	1.117	15.07	0.94	5.9
1818*	0.737	16.16	0.67	7.7	827*	1.129	14.80†	1.07	4.2
815*	0.763	15.89†	0.75	9.5	2463	1.145	14.60	1.61	4.4
2727	0.775	15.55	1.21	10.5	1006	1.153	15.00	1.50	5.1
2619	0.777	15.41	1.38	10.0	1335*	1.158	14.98†	1.23	4.6
2124*	0.783	15.72†	0.82	10.0	933	1.192	15.40	1.49	14.3
5954	0.788	15.79	0.56	8.0	1328*	1.200	14.46	1.27	4.6
1492*	0.799	15.42†	1.19	11.8	1333*	1.212	15.10†	1.37	6.5
2773	0.803	15.55	1.05	8.7	1342*	1.254	14.34†	0.67	5.7
2536	0.804	15.17	1.31	7.4	1005	1.272	14.69	1.46	15.4
2685	0.816	15.40	1.08	7.7	817*	1.276	14.01†	1.16	6.2
2790	0.820	16.01	0.90	11.8	1003	1.387	13.83	1.22	14.3
1400*	0.823	15.64†	0.82	6.5	934	1.450	14.75	1.50	11.1
2358	0.825	15.67	0.70	10.0	823*	1.504	14.44†	1.20	20.0
1855*	0.835	15.89†	0.76	11.8	855*	1.518	14.54†	1.40	10.5
2337	0.837	15.44	1.17	11.8	953	1.680	13.18	1.37	10.0
935	0.849	15.23	1.19	10.0	2447	2.074	13.16	1.01	5.3
2491	0.853	15.36	1.12	11.1					

* Variable in Small Magellanic Cloud.

† Magnitudes by A. S. Carlson; others by V. McK. Nail.

$$r\,(\text{p.i.}, M) = 0.41.$$

The near equality of these computed coefficients is to be expected since the period-luminosity relation for these 89 stars gives the well-known highly significant correlation,

$$r\,(M, \log P) = 0.91.$$

The scatter of points in figure 2 shows for this homogeneous material the strength and weakness of the so-called Hertzsprung effect. From the plot, we would say that from periods of five days to periods of fifteen days the steepness tends to decrease measurably. There are, however, several non-conforming stars with periods between eight and twelve days. This is also the interval where double maxima appear, but the double-maxima stars (indicated by arrows) show no trend in steepness with period.

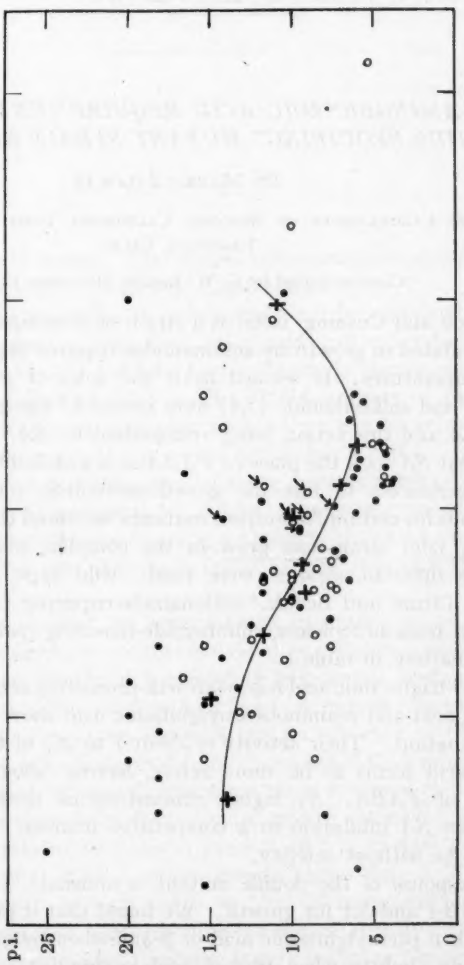

FIGURE 2
The relation of pulse index (ordinates) to logarithm of period.

A plot of pulse index against photographic amplitude shows no interdependence of these quantities; the correlation coefficient,

$$r \text{ (p.i., amp.)} = 0.02,$$

is in agreement with that conclusion.

[1] Current estimates are 23 and 26 kpc, respectively.
[2] Shapley, McKibben and Mohr, these PROCEEDINGS, **26**, 326 (1940).
[3] Gaposchkin, S., *Ibid.*, **24**, 1 (1938).
[4] Gaposchkin, C. P., *Ast. Jour.*, **52**, 218–226 (1947).

---

*THE p-AMINOBENZOIC ACID REQUIREMENT OF THE "SULFONAMIDE-REQUIRING" MUTANT STRAIN OF NEUROSPORA**

BY MARKO ZALOKAR

KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIF.

Communicated by G. W. Beadle, December 13, 1947

Emerson and Cushing[1] isolated a strain of *Neurospora* which instead of being inhibited in growth by sulfonamides requires the drug for growth at high temperatures. It seemed as if the roles of *p*-aminobenzoic acid (*PABA*) and sulfanilamide (*SA*) were reversed,[2] the mold being poisoned by *PABA* and this action being antagonized by *SA*. The evidence suggested that *SA* took the place of *PABA* as a metabolite.

In experiments to test the growth-promoting activity of folic acid compounds for certain *Neurospora* mutants, we found that the sulfonamide-requiring (*sfo*) strain can grow in the complete absence of *SA*. The following different strains[3] were used: wild type, *p*-aminobenzoicless (*pab*) of Tatum and Beadle,[4] sulfonamide-requiring (*sfo*) and the double mutant *p*-aminobenzoicless, sulfonamide-requiring (*pab, sfo*). The results are summarized in table 1.

Pteroyl-triglutamic acid has no growth-promoting action at all. Pteroylglutamic acid and *p*-aminobenzoylglutamic acid show growth activity for the *pab* mutant. Their activity is about 1 to 2% of the *PABA* activity. Pteroic acid seems to be more active, having about 5 to 10% of the activity of *PABA*. At higher concentrations these three substances antagonize *SA* inhibition in a competitive manner. Pteridine was also found to be without activity.

The response of the double mutant is unusual. It requires normally both *PABA* and *SA* for growth. We found that it grows in the absence of *SA* when pteroylglutamic acid or *p*-aminobenzoylglutamic acid is supplied, only slightly when pteroic acid is supplied. In the presence of 10^{-4} *M SA* the growth is poor with the first two substances, but good with the third one. In the presence of 10^{-7} *M PABA* there is no growth with

TABLE 1

GROWTH OF DIFFERENT STRAINS OF *Neurospora* ON FOLIC ACID AND RELATED COMPOUNDS; +++ GOOD GROWTH, ++ FAIR GROWTH, + POOR GROWTH, − NO GROWTH

		COMPOUNDS ADDED ($10^{-8}M$)					
STRAIN	MEDIUM	NONE	PETEROYL-TRIGLUTAMIC ACID	PTEROYL-GLUTAMIC ACID	PTEROIC ACID	*p*-AMINO-BENZOYL-GLUTAMIC ACID	*p*-AMINO-BENZOIC ACID
Wild type	Minimal	+++	+++	+++	+++	+++	+++
	10^{-2} *M SA*	−	−	−	−	−	++
pab	Minimal	−	−	++	+++	+++	+++
pab, sfo	Minimal	−	−	++	+	++	−
	10^{-4} *M SA*	−	−	+	+++	++	+++
	10^{-7} *M PABA*	−	−	−	−	−	−
sfo	Minimal	−	−	−	−	−	−
	10^{-4} *M SA*	+++	+++	+++	+++	+++	++

any of the substances tested. The simple sulfonamide-requiring mutant does not grow on any of the folic acid compounds.

The action of all three folic acid compounds can be explained by their free arylamine content. Crystalline folic acid (pteroylglutamic acid) contains about 0.5 to 1% *PABA*, *p*-aminobenzoylglutamic acid about 0.1% and pteroic acid about 6%.[5, 6] The *PABA* activity of these compounds is in some instances somewhat higher than can be accounted for by the free arylamine content; it seems likely that a certain amount of cleavage to *PABA* occurs during autoclaving or during growth of the mold. From their inability to substitute for *PABA*, it may be concluded either that folic acid and its derivatives are not products of *PABA* metabolism in *Neurospora*, or else that they have failed to penetrate into the cell in these experiments.

As for the double mutant, we assumed that it grows only at very low *PABA* concentrations, such as those provided as impurities in folic acid. This hypothesis was confirmed by growing the double mutant at low *PABA* concentrations. It can grow well in a very narrow range of *PABA* concentrations (Fig. 1). If we compare the growth rate of the double mutant with that of *pab*, we see that from the lowest *PABA* concentration up both curves ascend at exactly the same slope to about 2×10^{-8} to 5×10^{-8} *M*, where they reach the optimum growth velocity. The curve for the double mutant drops rapidly above 5×10^{-8} *M* to show a typically *SA*-requiring growth at *PABA* concentrations higher than 2×10^{-7} *M*. The growth between 5×10^{-8} and 2×10^{-7} *M* is typically adaptive: the velocity improves with time, reaching the normal rate at the end. At higher concentrations of *PABA* the curve is very irregular, owing to differences in the degree of adaptation to the absence of *SA*, and to "reversions" caused by mutation.[7]

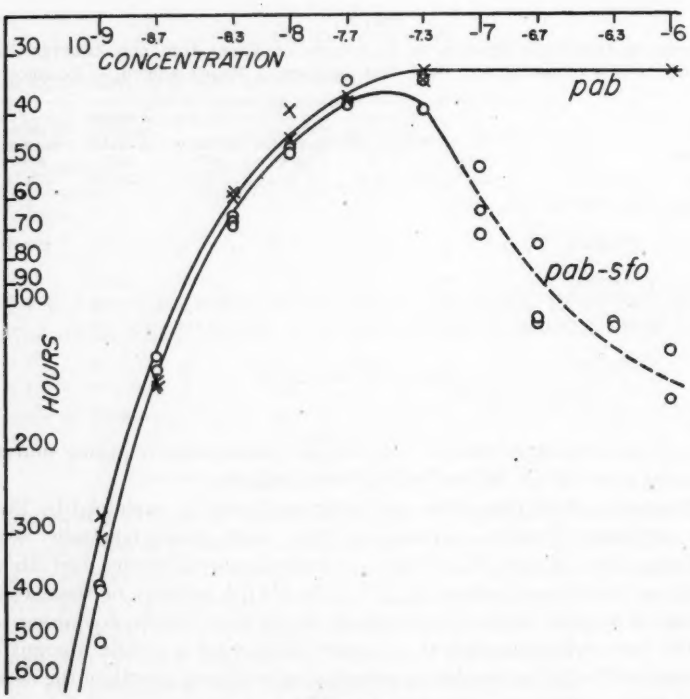

FIGURE 1

Growth of the *p*-aminobenzoicless mutant (*-x-x-*) and the double mutant *p*-aminobenzoicless, sulfonamide-requiring (*-o-o-*) on different concentrations of *p*-aminobenzoic acid (experiments Z — 90, Z — 101). Abscissa: concentration of *PABA* in moles; ordinata: time necessary for mycelium to reach 150 mm. in growth tubes.

Looking back to the action of the folic acid compounds, the observed results can be completely explained by the amounts of free *PABA* they contain. These amounts, as deduced from growth of the double mutant, are in agreement with the values obtained by assay with the *p*-aminobenzoicless mutant.

There is only one simple explanation of these facts: in the presence of the gene conditioning the *SA* requirement, *PABA* is poisonous in concentrations which are harmless to the wild type. As the inhibition by *PABA* (at concentrations less than 10^{-3} *M*) is completely relieved by *SA* in a competitive manner, we can draw a scheme comparable to the scheme of *PABA* action by Kohn and Harris.[8] They assumed that *PABA* is involved in the synthesis of substances x_1, x_2, $\ldots x_n$ which are necessary

for growth. *SA* competes with *PABA* and inhibits these syntheses. In our case the *sfo* gene would direct a reaction whereby *PABA* produces some inhibitory substance, *y*. This reaction must require a somewhat higher *PABA* concentration than is needed for normal growth and therefore is not apparent below a concentration of 10^{-7} *M PABA*. Further, the reaction must be very sensitive to *SA* inhibition and be stopped before *SA* inhibits growth by affecting vital reactions (leading to x_1, x_2, etc.). Search is under way for the particular reaction involved.

The sulfonamide-requiring mutant, in the absence of the *pab* gene, obviously produces more *PABA* than corresponds to a concentration of 10^{-7} *M*, so that it poisons itself and requires *SA* for growth as a detoxicant. Hence we have to drop the idea that *SA* is utilized as a metabolite.

Summary.—1. *Neurospora* cannot use pteroylglutamic acid, pteroic acid or *p*-aminobenzoylglutamic acid to replace *p*-aminobenzoic acid.

2. These folic acid compounds contain a certain amount of free *PABA*, which explains a positive growth response of the *p*-aminobenzoicless mutant and the double mutant *p*-aminobenzoicless, sulfonamide-requiring.

3. The double mutant can grow well on low concentrations of *PABA* (10^{-8} to 10^{-7} *M*), but is poisoned by greater concentrations and requires *SA* as a detoxicant.

4. The sulfonamide-requiring strain must produce more than the tolerated amount of *PABA* and thus inhibits itself.

Acknowledgment.—The author wishes to express his indebtedness to Dr. Sterling Emerson for suggesting the problem and for offering advice and help throughout the course of the work.

* Representing work supported in part by a grant from the American Cancer Society upon the recommendation of the Committee on Growth of the National Research Council.

[1] Emerson, S., and Cushing, J. E., "Altered Sulfonamide Antagonism in *Neurospora*," *Federation Proc.*, **5**, 379–389 (1946).

[2] Emerson, S., "Growth Responses of a Sulfonamide-Requiring Mutant Strain of *Neurospora*," *J. Bact.*, **54**, 195–207 (1947).

[3] The particular strains used are: wild type E5256A, *pab* 1633 reisolated as Z35-I-8, *sfo* E15172 reisolated as E16014A and *pab*, *sfo* 1633 and E15172 combined in strain E19335.

[4] Tatum, E. L., and Beadle, G. W., "Genetic Control of Biochemical Reactions in *Neurospora:* An Aminobenzoicless Mutant," *Proc. Nat. Acad. Sci.*, **28**, 234–243 (1942).

[5] Williams, R. D., "The Comparative Antisulfanilamide Activity of *p*-Aminobenzoyl-1(+)-glutamic Acid and *p*-Aminobenzoic Acid," *J. Biol. Chem.*, **156**, 85–89 (1944).

[6] Lampen, J. O., and Jones, M. J., "The Antagonism of Sulfonamide Inhibition of Certain Lactobacilli and Enterococci by Pteroylglutamic Acid and Related Compounds," *Ibid.*, **166**, 435–448 (1946).

[7] Emerson, S., "A Physiological Basis for Some Supressor Mutations and Possibly for One Gene Heterosis," *Proc. Nat. Acad. Sci.* (in press).

[8] Kohn, H. I., and Harris, J. S., "On the Mode of Action of Sulfonamides. I. Action on *Escherichia coli*," *J. Pharm. Exp. Therap.*, **73**, 343–361 (1941). Harris, J. S., and

Kohn, H. I., "On the Mode of Action of Sulfonamides. II. The Specific Antagonism Between Methionine and the Sulfonamides in *Escherichia coli*," *Ibid.*, **73**, 383–400 (1941).

DIRECTED FERTILIZATION IN MAIZE

By Herschel Roman

California Institute of Technology* and the University of Washington†

Communicated by L. J. Stadler, December 15, 1947

Ordinarily the two gametes in the maize pollen grain are genetically identical. However, a plant carrying an A-B interchange produces a pollen grain in which the gametes are not alike.[1] This paper deals with the rôle of the dissimilar gametes in fertilization.

An A-B interchange is one between the supernumerary B-type chromosome[2] and a member of the basic complement (A-type chromosome). The behavior of the interchange chromosomes during the development of the pollen grain may be illustrated by the case of TB-4a. In this interchange, the distal seven-eighths or so of the short arm of chromosome 4 was transferred to a segment of a B-type bearing the centromere of the latter. The resulting chromosome is designated the B^4 chromosome. The other interchange chromosome (4^B) contains the rest of chromosome 4 and most or all of the distal heterochromatic segment of the B-type.

The 4^B chromosome is orthodox in its behavior and is found regularly in each gamete of the pollen grain. The B^4 chromosome also follows the normal pattern until the division of the generative nucleus is reached. In this division the B^4 chromosome frequently undergoes non-disjunction so that one of the gametes receives two B^4 chromosomes and the other receives none. A second type of pollen grain is the result of normal disjunction; both gametes of the pollen grain are identical, each containing one B^4 chromosome.

Three classes of seed are expected when a normal seed parent is crossed with a pollen parent homozygous for TB-4a. Two of these are obtained from fertilization involving the first type of pollen grain. If the gamete that is deficient for the B^4 chromosome fertilizes the egg and its partner unites with the polar nuclei, a seed with a deficient embryo and a hyperploid endosperm will result (Class I). If the gametes exchange their respective rôles in fertilization, the seed will have an embryo hyperploid for the B^4 chromosome and a deficient endosperm (Class II). A third kind of seed is produced when the second pollen type is involved; in this case, the embryo is heterozygous for the interchange and the endosperm also carries a single B^4 chromosome (Class III).

The three classes of seed may be readily identified through the use of the sugary endosperm gene, *su*. The B^4 chromosome carries the dominant allele *Su*. If the seed parent is homozygous for *su*, the three kinds of seed would have, respectively:

Class I: A non-sugary, hyperploid endosperm and a deficient embryo, hemizygous for *su;*

Class II: A sugary, deficient endosperm and an embryo of composition *su Su Su;* and

Class III: A non-sugary, euploid endosperm with a euploid embryo of composition *su Su*.

If fertilization involving the first type of pollen grain occurs in either direction with equal frequency (i.e., random fertilization), seed classes I

TABLE 1

RESULTS OF CROSSES BETWEEN NORMAL SEED PARENTS HOMOZYGOUS FOR *su* AND POLLEN PARENTS HOMOZYGOUS FOR TB-4a AND *Su*

CROSS	SUGARY SEEDS	TOTAL SEEDS	% SUGARY	±S.E.
GC-1 × 25-2	162	271	59.8	3.0
2 × 25-2	175	267	65.5	2.9
3 × 25-2	141	217	65.0	3.2
4 × 25-2	158	229	69.0	3.1
5 × 25-2	106	218	48.6	3.4
107-3 × 25-2	164	274	59.9	3.0
107-4 × 25-2	136	262	51.9	3.1
2268-3 × 25-2	130	210	61.9	3.4
2269-2 × 25-2	226	410	55.1	2.5
2269-7 × 25-2	314	504	62.3	2.2
108-28 × 50-13	76	132	57.6	4.3
273-2 × 50-13	124	216	57.4	3.4
263-10 × 50-14	169	303	55.8	2.9
277-4 × 50-14	74	118	62.7	4.5
Totals	2155	3631	59.4	0.8

and II should be equal in number assuming they are equally viable. The number of seeds with a sugary endosperm (Class II) would not be expected, on this basis, to exceed the number of non-sugary seeds. Actually, the sugary seeds should be in the minority since the non-sugary seeds are expected to include not only those of Class I but also those of Class III, the latter derived from fertilization involving the second type of pollen grain. The results of crosses published previously[1] gave an average value of approximately 50% for the proportion of sugary seeds and this was interpreted, assuming random fertilization, as an indication of a very low rate of normal disjunction. There were two crosses, however, in this first

group that gave what appeared to be a significantly higher proportion of sugary seeds. Subsequent crosses involving three different but closely related pollen parents, cytologically identified as homozygous for TB-4a, have given a segregation of sugary seeds clearly in excess of 50% (table 1). It should be noted that these data are not selected but represent the total harvest of crosses involving these parents.

The extent to which the sugary seeds outnumber their reciprocal counterparts (Class I) cannot be determined directly from the data in table 1, since the non-sugary seeds that have a hyperploid endosperm are indistinguishable from those with a euploid endosperm (Class III). The two types of non-sugary seeds may, however, be separated on the basis of the difference between their embryos. This may be accomplished, in the plants grown from these seeds, by a direct cytological examination of the microsporocytes, by an examination of the mature pollen for the percentage of aborted grains, or on the basis of sugary ratios in crosses with a *su su* parent. The identification methods employed will be described in detail elsewhere.

TABLE 2

EMBRYO COMPOSITION OF NON-SUGARY AND SUGARY SEEDS OF CROSSES GIVEN IN TABLE 1. SEE TEXT FOR FURTHER EXPLANATION

	NON-SUGARY SEEDS			SUGARY SEEDS		
CROSS	− (CLASS I)	+ (CLASS III)	++	−	+	++ (CLASS II)
GC-4 × 25-2	25	28	1	1	1	131
GC-5 × 25-2	40	32	2	0	0	87
2269-2 × 25-2	44	16	2	0	1	83
2269-7 × 25-2	43	48	0	0	0	81
Totals	152	124	5	1	2	382

Table 2 shows the distribution of embryo types found in non-sugary seeds. Also included for comparison are the types contained in sugary seeds. The symbols −, + and ++ refer, respectively, to: (1) an embryo deficient for the B^4 chromosome, (2) a heterozygous embryo, carrying one B^4 chromosome and (3) a hyperploid embryo, with two B^4 chromosomes. Among the non-sugary seeds, only about half contain the deficient embryo indicative of Class I.

The data in table 2 also reaffirm the conclusion reached from previous evidence that a plant homozygous for TB-4a produces two types of functional pollen grains. With only eight exceptions among 681 seeds tested, the seeds fall into the three classes described above. The exceptions would be expected normally, as a consequence of heterofertilization,[3] in which the gametes of two different pollen grains unite with the egg and polar nuclei of the same embryo sac.

If we assume from the data in table 2 that half of the non-sugary seeds belong to Class III and apply this correction to the data in table 1, the

preponderance of Class II seeds over Class I becomes even more evident. The cross 2269-7 × 25-2 gave 314 sugary seeds and 190 non-sugary seeds; when the correction is applied, the ratio of Class II to Class I is 314 to 95, or roughly 3:1. In the cross GC-5 × 25-2, which gave a relatively low percentage of sugary seeds, the ratio is 106 to 56, or approximately 2:1. If we take 60% as an average value for the frequency of sugary seeds in all of the crosses of table 1, the proportion of Class I:Class II:Class III becomes, using the same correction, 1:3:1. Thus, four-fifths of the progeny were obtained from fertilizations involving pollen grains in which non-disjunction had occurred and in three-fourths of these, the deficient gamete united with the polar nuclei and the hyperploid gamete with the egg.

It was expected, assuming random fertilization, that the Class I and Class II seeds would occur in equal numbers. The low frequency of Class I seed brings up the question of the relative viability of this type. Are these seeds being lost in development? If this were so, it should result in a detectable amount of sterility on the ear since the loss would comprise two-thirds of this class or two-sevenths of the total population. Actually, several of the ears in table 1, including those which gave the most disproportionate ratios, had full sets of seed. It seems likely, therefore, that the low proportion of Class I seeds is evidence that fertilization is not random; that the union of the egg with the hyperploid gamete and the concomitant fusion of the polar nuclei with the deficient gamete is preferred to the reverse order of fertilization, in which the deficient gamete unites with the egg and the hyperploid gamete with the polar nuclei.

Similar results have been obtained with TB-9b, an interchange involving chromosome 9 of the basic set. The break in chromosome 9 occurred about halfway along the short arm, between *C* and *Wx;* the B-type chromosome was broken within the distal heterochromatic segment. In the cross to be described, the B^9 chromosome carried *C* (for anthocyanin-colored endosperm) and 9^B chromosome carried *Wx* (for starchy endosperm). The TB-9b pollen parents used in the cross were homozygous for the 9^B chromosome and hemizygous for the B^9 chromosome (with respect to chromosome 9, their constitution was 9^B 9^B B^9). Such a plant is equivalent, as regards male transmission, to the homozygote (with two B^9 chromosomes), since only the microspores possessing both a B^9 and a 9^B chromosome develop into functional pollen grains. The microspores that are deficient for the B^9 chromosome produce aborted pollen grains with a distinctly reduced starch content.

The seed parents used in the cross were cytologically normal and homozygous for *c* and *wx*. In the TB-9b pollen parent, the B^9 chromosome undergoes non-disjunction in the division of the generative nucleus. A pollen grain is thus produced in which one gamete has two B^9 chromosomes

and the other has none (and therefore lacks *C*). The union of the hyperploid gamete with the polar nuclei and the deficient gamete with the egg would give a kernel with a hyperploid, colored endosperm and a deficient embryo (analogous to the Class I seed obtained from the crosses involving TB-4a). The reciprocal order of fertilization would produce a seed with a deficient, colorless endosperm and a hyperploid embryo (Class II). The results of six crosses, involving two TB-9b parents are given in table 3. Here, as with TB-4a, the seeds with a deficient endosperm are clearly in the majority. With rare exception, the seeds of both classes were starchy, an indication of the regularity of disjunction of the 9^B chromosome in the division of the generative nucleus.

TABLE 3

SEE TEXT FOR DESCRIPTION OF CROSS

CROSS	SEEDS WITH COLORLESS ENDOSPERM	TOTAL SEEDS	% COLORLESS	±S.E.
119-4 × 96-8	96	148	64.9	3.9
119-11 × 96-8	235	356	66.0	2.5
2264-7 × 96-8	198	334	59.3	2.7
119-5 × 96-15	214	305	70.2	2.6
119-10 × 96-15	150	249	60.2	3.1
2264-11 × 96-15	220	353	62.3	2.6

Scutellum color was used as an indicator of the presence of the B^9 chromosome in the embryos of the colored and colorless seeds. *C*, in addition to certain other known genes, is required for anthocyanin color in the scutellum.[4] Two of the crosses gave an appropriate complementary genotype and the seeds obtained from these were classified for scutellum color in relation to endosperm color. It was anticipated that the Class I seeds should have a colorless embryo and the Class II seeds should have a colored embryo. The results given in table 4 show that over 97% of the

TABLE 4

RELATION OF SCUTELLUM COLOR TO ENDOSPERM COLOR IN TWO CROSSES OF TABLE 3

CROSS	COLORED ENDOSPERM*		COLORLESS ENDOSPERM†	
	COLORED SCUTELLUM	COLORLESS SCUTELLUM	COLORED SCUTELLUM	COLORLESS SCUTELLUM
119-4 × 96-8	0	52	95	1
119-11 × 96-8	9‡	108	232	2
Totals	9	160	327	3

* Three others lacked an embryo and one other was also unclassifiable for scutellum color.

† One other lacked an embryo.

‡ One had a colorless endosperm sector.

seeds fall into these two categories and that the Class II seed outnumbers its reciprocal counterpart by approximately 2:1.

Twelve other seeds, not of Class I nor II, were also obtained. Three had a colorless endosperm as well as a colorless embryo; they were therefore deficient in both endosperm and embryo for the B^9 chromosome. They may be accounted for as cases of heterofertilization in which deficient gametes from two different pollen grains were involved. The nine seeds with a colored endosperm and a colored embryo may also be the result of heterofertilization, this time involving two hyperploid gametes. Or they may represent fertilization involving a pollen grain in which both gametes contain one B^9 chromosome, as a consequence of normal disjunction of this chromosome (thus equivalent to the Class III seeds of table 2). The frequency of normal disjunction appears to be much lower for the B^9 chromosome than for the B^4 chromosome. Whether this is a characteristic difference between the two chromosomes or is an expression of a difference in genotype in the two parental types is a question for further investigation.

The evidence from the crosses with TB-4a and TB-9b may be summarized as follows. Plants carrying either of these interchanges produce, in the main, one type of pollen grain. This contains one gamete deficient for the aberrant chromosome (B^4 or B^9, as the case may be) and another gamete carrying the chromosome in duplicate. Fertilization involving the two gametes may proceed in either of two ways: (1) the deficient gamete may unite with the polar nuclei to form the primary endosperm nucleus and the hyperploid gamete fertilizes the egg or (2) the hyperploid gamete may unite with the polar nuclei and the deficient partner with the egg. The first fertilization pattern is much more prevalent than the second. In the case of TB-4a, it was estimated that the first occurred three times as frequently as the second; with TB-9b, it was twice as frequent. The preference of one pattern over the other will be referred to by the term "directed fertilization."

Two possible explanations that will account for directed fertilization are being tested. The first of these assumes a specific orientation of the deficient and the hyperploid gametes within the pollen grain or at some time prior to fertilization. According to this hypothesis one of the gametes would have a positional advantage over the other that would result in the preferential union of the hyperploid gamete with the egg and the deficient gamete with the polar nuclei. The other explanation assumes a physiological difference between the hyperploid and deficient gametes that results in directed fertilization.

The assumption that the gametes are oriented within the pollen grain presupposes that one of the oriented gametes receives both of the non-disjoining chromosomes more often than the other. Evidence for a directed distribution such as this has been demonstrated cytologically in other organisms in which mitotic non-disjunction occurs. In *Sciara co-*

prophila, the sex chromosome moves to a specific pole in the division of the secondary spermatocyte.[5] In rye, the supernumerary chromosome, which undergoes non-disjunction in the first division of the microspore, is found more often in the generative nucleus than in the vegetative nucleus.[6] The products of non-disjunction of the supernumerary chromosome in sorghum are directed to the generative pole in the exceptional division of the vegetative nucleus.[7]

It is doubtful that the deficiency or duplication of A-chromatin (the segment of chromosome 4 in the B^4 chromosome or chromosome 9 in the B^9 chromosome) in the gametes is an essential factor in determining directed fertilization. An indication of directed fertilization is also found in the results of a cross in which the pollen parent had two intact B-type chromosomes and a normal basic complement. When the seed parent was devoid of B-types, most of the progeny possessed either two B-types or none in their embryos. The progeny with two B-types were in a marked majority.[2, 8] This may be interpreted as follows: (1) the 2B parent produces a pollen grain in which one gamete has two B-types and the other has none; and (2) the 2B gamete has an advantage over its OB partner in the fertilization of the egg.

Directed fertilization of the egg by the 2B gamete is also of interest in connection with the problem of the survival of the B-type chromosome. This chromosome is subject to fragmentation and also to loss in meiosis.[2, 9] Yet it persists in a number of maize varieties and is particularly prevalent in certain sugary strains, notably the Black Mexican. The preferential fertilization of the egg by the gamete containing the B-types, after non-disjunction, would provide an excess of B-types to counteract the fragmentation and loss of this chromosome. The common occurrence of the B-type in certain varieties could then be accounted for if the mechanism of directed fertilization is more highly developed in these strains, as compared with others in which the B-type is absent or rarely found.

* Gosney Fellow. I am indebted to Dr. E. G. Anderson for his generous coöperation.

† Permanent address. A grant from the Agnes H. Anderson Research Fund for technical assistance is gratefully acknowledged.

[1] Roman, H., *Genetics*, **32**, 391–409 (1947).

[2] Randolph, L. F., *Ibid.*, **26**, 608–631 (1941).

[3] Sprague, G. F., *Ibid.*, **17**, 358–368 (1932).

[4] Sprague, G. F., *U. S. Dept. Agr. Tech. Bull.*, **292**, 1–43 (1932).

[5] Metz, C. W., *Am. Naturalist*, **72**, 485–520 (1938).

[6] Müntzing, A., *Hereditas*, **32**, 97–119 (1946).

[7] Darlington, C. D., and Thomas, P. T., *Proc. Roy. Soc. London*, **130**, 127–150 (1941).

[8] Supplementary data (unpublished) were kindly furnished by Dr. A. E. Longley. Among the examined progeny from seven OB × 2B crosses, 31 had 2B's and 13 had none.

[9] Darlington, C. D., and Upcott, M. B., *J. Genet.*, **41**, 275–296 (1941).

ABSTRACT ERGODIC THEOREMS*

BY W. F. EBERLEIN

THE INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

Communicated by J. von Neumann, December 17, 1947

1. The mean ergodic theorem of J. von Neumann has been extended by various authors—notably F. Riesz[1] and Yosida and Kakutani[2]—to an abstract theorem asserting the convergence to a fix point of the means $T_n x = (1/n + 1)\sum_0^n T^j x$, where T is a linear transformation in a Banach space E. Alaoglu and Birkhoff[3] replaced the iterates (T^j) by a semi-group G of linear transformations and showed that the convergence of certain general means of transforms of an element x of E is equivalent to the existence (and uniqueness) of a fix point y in the closed convex hull of the orbit of x under G. In this note we indicate how relaxation of the uniform boundedness and countability restrictions customarily imposed on G leads to a more general theory, which not only embraces a significantly wider domain of phenomena but subsums previous results in a sharper and more transparent form. Details and further developments will appear elsewhere.

2. Consider a Banach space E and a semi-group G of linear transformations T of E into itself. We denote by $G^* = [\sum a_j T_j | a_j \geqq 0, \sum a_j = 1, T_j \epsilon G]$ the family of transformations T^* consisting of all finite convex combinations of the elements T of G, by $\overline{G}^*$ the closure of G^* in the uniform topology of operators, by $0(x) = [T^*x | T^* \epsilon G^*]$ the orbit of x under G^*, and by $\overline{0}(x)$ the closure of $0(x)$ in E. G^* and $\overline{G}^*$ are clearly semi-groups of operators, and $0(x)$ and $\overline{0}(x)$ are convex sets. We may and shall assume that G contains the identity transformation.

Definition 1: G is an ergodic semi-group if it possesses at least one system of almost invariant integrals. By such a system we mean a family of transformations (T_α), indexed by the directed set (α), with the following properties:

I. T_α is a linear transformation of E into itself for every α.
II. For every x and all α $T_\alpha x \epsilon \overline{0}(x)$.
III. There exists an $M > 0$ such that $||T_\alpha|| \leq M$ for all α.
IV. For every x in E and all T in G
(a) $\lim_\alpha (TT_\alpha x - T_\alpha x) = 0$;
(b) $\lim_\alpha (T_\alpha Tx - T_\alpha x) = 0$.

If the limits IV hold uniformly for all x in any bounded set, we call the system (T_α) a uniform system. IV remains valid if T is replaced by any element U of $\overline{G}^*$.

3. The following examples provide some indication of the scope of the theory. Obvious generalizations will suggest themselves. It must be emphasized (Theorem 1 infra) that the essential phenomena depend on the existence of *at least one* system (T_α), rather than on the particular system exhibited.

Example 1: Standard Case. G consists of the iterates (T^n) of a linear transformation T and the T_α are the previously defined T_n under the natural ordering. I and II are trivially satisfied, and the usual condition that $||T^n|| \leq M$ for all n ensures the validity of III, and of IV in the uniform sense. However, as pointed out in examples by Hille[4] and Dunford,[5] uniform boundedness of the T^n is too restrictive: all one needs is a bound on the $||T_n||$ and $\lim T^n/n = 0$ in an appropriate sense.

Example 2: Almost Periodic Functions. E is the Banach space of complex-valued bounded continuous functions $f(x)$ $(-\infty < x < \infty)$ with $||f|| = \sup |f(x)|$. G is the group of translations $f(x) \to f(x + a)$, whose only fix points are the constant functions. One may take as the system (T_α) the Bohr means

$$T_\alpha f = \frac{1}{\alpha} \int_0^\alpha f(x + t)dt \qquad (\alpha > 0)$$

under the natural ordering.

Example 3: Bounded Abelian Semi-Group. If G is Abelian and if $||T|| \leq M$ for all T in G, then the elements of G^* form a uniform system of almost invariant integrals under the ordering: $U \supset V$ if and only if there exists W in G^* such that $U = WV$. (That this ordering has the required composition property follows from the fact that $UV = VU$ is a common successor of U and V.) Clearly this example includes Example 2 and the bounded case of Example 1.

Example 4: Summability of Fourier Series. E is the space C of continuous real-valued functions of period 2π, with $||f|| = \sup |f(x)|$. G consists of the transformations $U_n f = f - S_n$, where $S_n(x)$ is the sum of the first $2n + 1$ terms of the formal trigonometrical series corresponding to f. The identity $U_n U_m = U_\nu$, where $\nu = \max.(m, n)$, implies that G is an Abelian semi-group, whose only fix point is the null function. The $||U_n||$ are unbounded—if bounded the U_n themselves would form a system of almost invariant integrals for G; and this hypothesis would in turn imply the uniform convergence of the Fourier Series of every continuous function, which is false. However, the $(C, 1)$ means $T_n = (1/n + 1)\sum_0^n U_j$ do form a system of almost invariant integrals, the existence of the bound in III following from the uniform $L_1(-\pi, \pi)$ bound of the Fejér kernel.

The above examples are all of the Abelian type. The more difficult non-commutative case is discussed in § 5.

4. We come now to the first key theorem and the definition of an ergodic element that it inspires.

THEOREM 1. *If G is ergodic, x an element of E, and (T_α) ANY system of almost invariant integrals, then the following conditions on an element y in E are equivalent:*

(1) $y \in \overline{0}(x)$, *and* $Ty = y$ *for all* $T \in G$;
(2) $y = \lim_\alpha T_\alpha x$;
(3) $y = \lim_\alpha T_\alpha x$ *weakly;*
(4) y *is a weak cluster point of* $(T_\alpha x)$.

Definition 2: If G is ergodic, x is ergodic with (unique) limit fix point y if and only if there exists a $y = T_\infty x$ satisfying any of the conditions (1) to (4) above.

THEOREM 2. *If G is ergodic, the ergodic elements of E constitute a closed invariant subspace Γ. The transformation $T_\infty x = y = \lim_\alpha T_\alpha x$ is a linear transformation of the space Γ into itself, and $||T_\infty||_\Gamma \leq M$. Moreover, on Γ we have $T_\infty = T_\infty^2 = T_\infty U = UT_\infty$ for every U in $\overline{G}^*$.*

The implications (2) → (3) → (4) of Theorem 1 are trivial. In verifying the crucial implication (4) → (1) we employ the Mazur theorem[6] to obtain $y \in \overline{0}(x)$ and then show that $f(Ty - y) = 0$ for all T in G and all f in E^*. IV (a) is required only to establish (4) → (1), IV (b) only in proving (1) → (2).

Condition (4) is the most useful condition for ergodicity. The required y exists automatically when the set $(T_\alpha x)$ or $[Tx | T \in G]$ is (conditionally) weakly compact. From this remark follow the standard mean ergodic theorems (Example 1) and the existence and uniqueness of the mean for generalized almost periodic functions defined on Abelian groups (Examples 2 and 3). Proof of the Fejér theorem (Example 4) proceeds thus: Since $\lim_n T_n f = 0$ in $L_2(-\pi, \pi)$, there exists a subsequence $T_{n_i} f$ converging to 0 almost everywhere. The Lebesgue convergence theorem then implies that $\lim_n T_{n_i} f = 0$ weakly in C, whence $\lim_n T_n f = 0$ strongly in C.

Conditions on the operators T of G sufficient to ensure ergodicity are of interest. Yosida and Kakutani found the special (bounded) case of Example 1 in which T is (weakly) quasi-completely continuous fundamental in their abstract treatment of Markoff processes.[2] [A transformation T is called (weakly) quasi-completely continuous if there exists an integer m and a (weakly) completely continuous transformation V such that $||T^m - V|| < 1$.]

THEOREM 3. *If G is (ergodic) uniformly ergodic and if some T in $\overline{G}^*$ is (weakly) quasi-completely continuous, then every x in E is ergodic.* $T_\infty = \lim T_\alpha$ *(strongly) uniformly, and the projection T_∞ is (weakly) completely*

continuous. The set $T_\alpha(E)$ of fix points is a (reflexive) finite dimensional subspace of E.

Not only is this theorem a considerable generalization of the ergodic theorems of Yosida and Kakutani, but our proof is much simpler. The principal tool is the identity

$$T_\alpha = (I - D)^{-1}VT_\alpha + (I - D)^{-1}(T_\alpha - T^mT_\alpha),$$

where $D = T^m - V$.

5. In the general non-Abelian case the existence of a system of almost invariant integrals and the definition itself of an ergodic element are obscure. If the T in G are uniformly bounded, however, the ordering of G^* in Example 3, although no longer of Moore-Smith type, leads naturally to the Alaoglu-Birkhoff definition[7] of ergodicity:

Definition 3: If G is bounded, an element x of E is ergodic if and only if the means T^*x converge to a (unique fix) point y.

(When G is both bounded and ergodic, definitions 2 and 3 coincide.) The ergodic elements again form a closed invariant subspace Γ. It is clear that a necessary condition for ergodicity of x is that $\overline{0}(x)$ contain a unique fix point; that $\overline{0}(x)$ contain a unique fix point for each element x of an invariant set is a sufficient condition. Algebraic restrictions on the space E or the semi-group G may reduce the problem to establishing only the existence of a fix point y in $\overline{0}(x)$.

THEOREM 4: *Let G_1, G_2 be bounded semi-groups of operators such that $UV = VU$ for every U in G_1 and V in G_2, and let every fix point of G_1 be a fix point of G_2 and conversely. Then*

(1) The fix points of $\overline{0}_1(x)$ and $\overline{0}_2(x)$ reduce to a unique common fix point.

(2) If both $\overline{0}_1(x)$ and $\overline{0}_2(x)$ contain a fix point y for every x in an invariant set Γ, then every x of Γ is ergodic with respect to G_1, G_2, and G_1G_2 with the limit fix point y.

When $E = \Gamma$ is the set of complex-valued almost periodic functions—in the sense of von Neumann[8]—defined on an arbitrary group, one takes G_1 as the group of left translations, G_2 as the group of right translations. A combinatorial lemma of Maak[9] permits the construction of a system (T_α) satisfying I, II, III and IV (a)—sufficient to ensure the existence of a fix point in every $\overline{0}_1(x)$ or $\overline{0}_2(x)$. Theorem 4 then yields the existence of the mean in the form of an ergodic theorem. The existence of the invariant integral mean and Haar measure in compact groups is, of course, a special case. It is characteristic of this approach that the uniqueness theorem becomes an automatic consequence of Theorem 1.

* Prepared under the sponsorship of the Office of Naval Research.

[1] Riesz, F., *J. Lond. Math. Soc.*, **13**, 274–278 (1938).

[2] Yosida, K., and Kakutani, S., *Ann. Math.*, **42**, 188–228 (1941).

[3] Alaoglu, L., and Birkhoff, G., *Ibid.*, **41**, 293–309 (1940).

[4] Hille, E., *Trans. Am. Math. Soc.*, **57**, 246–269 (1945).

[5] Dunford, N., *Ibid.*, **54**, 185–217 (1943).

[6] For the requisite properties of the weak topology see Eberlein, W. F., these PROCEEDINGS, **33**, 51–53 (1947).

[7] For this definition and proofs of some of the immediately following remarks see Alaoglu, L., and Birkhoff, G., *loc. cit.*

[8] Neumann, J. von, *Trans. Am. Math. Soc.*, **36**, 445–492 (1934).

[9] Maak, W., *Hamburg Abhandlungen*, **11**, 240–244 (1935).

---

*CLOSED SURFACES WITHOUT CONJUGATE POINTS**

BY EBERHARD HOPF

N.Y.U. INSTITUTE FOR MATHEMATICS AND MECHANICS

Communicated by Marston Morse, December 8, 1947

The present note contains the proof of the following

THEOREM. *Let S be a closed surface of class C''' in the sense of Riemannian geometry. If no geodesically conjugate points exist on S the total curvature of S must be negative or zero. In the latter case the Gaussian curvature must vanish everywhere on S.*

From the second part of the theorem and from known facts one infers: if S is the topological image of the torus or of the Klein bottle and if F contains no geodesically conjugate points S must be the one-to-one and isometric image of a Euclidean model of such a surface.

This second part of the theorem also forms the subject of a recent paper by Morse and Hedlund.[1] These authors prove the theorem under an additional hypothesis about S (non-existence of focal points on S) and raise the question if the theorem holds true without this restriction.

The idea underlying the proof for the first part of the theorem had been outlined in a previous paper by the author.[2] It was found that only little additional remarks are required to make the proof cover the entire theorem. In the present note the proof will be developed ab ovo.

Proof of the Theorem. Consider the Jacobi differential equation of normal variation

$$y''(s) + K(s)y(s) = 0 \qquad (1)$$

(K = curvature) along an arbitrary oriented geodesic on S. The direction of increasing arc length is always chosen as to coincide with the direction of the geodesic. The solutions of (1) are well known to exist on the whole axis of s. Non-existence of conjugate points means: any non-trivial solution of (1) possesses at most one zero, any two non-identical solutions intersect at most once. Let

$$y(s;\ a,\ b)$$

be the (according to this intersection property unique) solution satisfying

$$y(a;\ a,\ b) = 1, \qquad y(b;\ a,\ b) = 0. \tag{2}$$

These functions satisfy the identity

$$y(s:\ a,\ b) = y(\alpha;\ a,\ b)y(s;\ \alpha,\ \beta) + y(\beta;\ a,\ b)y(s;\ \beta,\ \alpha). \tag{3}$$

Both sides represent solutions of (1) which, according to (2), intersect at $s = \alpha,\ \beta$ and which therefore coincide. In the special case $\alpha = a'$ and $\beta = b$ (3) becomes

$$y(s;\ a,\ b) = y(a';\ a,\ b)y(s;\ a',\ b). \tag{3'}$$

From (2) and from the intersection property we infer that

$$y(s;\ a,\ b') > 0 \text{ for } s < b' \text{ and } a < b'. \tag{4}$$

The two solutions $y(s;\ a,\ b)$ and $y(s;\ a,\ b')$, $b < b'$, intersect at $s = a$ but nowhere else. Hence, on account of (4),

$$y(s;\ a,\ b) \geqq y(s;\ a,\ b') \text{ for } s \leqq a < b < b'. \tag{4'}$$

(4) and (4') imply the existence of the limit

$$y(s;\ a) = \lim_{b=+\infty} y(s;\ a,\ b) \tag{5}$$

at any $s \leqq a$.

If in (3) $\alpha,\ \beta$ are both chosen less than a it becomes obvious that (5) exists at any s and that $y(s;\ a)$ is a solution of (1). It is also inferred that

$$y'(s;\ a) = \lim_{b=+\infty} y'(s;\ a,\ b) \tag{5'}$$

holds at any s. From (2) and (4) we get

$$y(a;\ a) = 1,\ y(s;\ a) \geqq 0$$

for any s. As $y(s;\ a)$ is a solution of (1) we even have $y > 0$ everywhere. The function

$$u(s) = \frac{y'(s;\ a)}{y(s;\ a)}$$

which, according to (3'), does not depend on a is thus continuous at every s. It is a solution of the Riccati equation

$$u'(s) + u^2(s) + K(s) = 0. \tag{6}$$

In this manner, a perfectly well-defined function u is obtained on any oriented geodesic of S. The value of this function is perfectly well de-

termined at every point of the geodesic and is obviously independent of the choice of the point where s is counted from. u is an everywhere continuous solution of (6) along the geodesic.

Let us note for later use that $u(s)$ is the limit of

$$\frac{y'(s;\ a,\ b)}{y\ (s;\ a,\ b)}$$

as $b \to \infty$. This fraction equals

$$\frac{Y'(s;\ b)}{Y\ (s;\ b)}$$

where $Y(s;\ b)$ is the solution of (1) satisfying $Y = 0$, $Y' = 1$ at $s = b$, and thus we may write

$$u(s) = \lim_{n=\infty} \frac{Y'_s(s;\ s+n)}{Y(s;\ s+n)}. \tag{7}$$

We next prove the boundedness of u for all geodesics on S. From the closedness of S one infers that

$$K > -A^2 \tag{8}$$

on S with a suitable constant $A > 0$. A Sturmian argument will show that

$$|u| \leqq A. \tag{9}$$

Let

$$z(s;\ a,\ b) = \frac{\sinh A(b-s)}{\sinh A(b-a)}$$

be the solution of

$$z''(s) - A^2 z(s) = 0 \tag{10}$$

satisfying

$$z(a;\ a,\ b) = 1, \qquad z(b;\ a,\ b) = 0.$$

From (1) and (10) we obtain

$$(zy' - yz')' = -(K + A^2)zy. \tag{11}$$

For $y = y(s;\ a,\ b)$ and $z = z(s;\ a,\ b)$, $a < b$, z and y are both positive for $s < b$. As $zy' - yz' = 0$ for $s = b$ we infer that $zy' - yz' > 0$ for $s < b$, i.e., that

$$u(s;\ b) = \frac{y'(s;\ a,\ b)}{y(s;\ a,\ b)} > \frac{z'(s;\ a,\ b)}{z(s;\ a,\ b)}$$

and therefore, on passing to the limit $b \to +\infty$, that

$$u(s) \geqq -A.$$

On the other hand, $y = y(s;\ a)$ and $z = z(s;\ a, c)$, $c < a$, are both positive for $s > c$, and it follows from (11) that

$$(zy' - yz')_s < (zy' - yz')_c = -y(c)z'(c) < 0 \text{ for } s > c.$$

We therefore have, for $s > c$,

$$u(s) = \frac{y'(s;\ a)}{y(s;\ a)} < \frac{z'(s;\ a, c)}{z(s;\ a, c)}$$

and finally, on letting $c \to -\infty$, $u \leqq A$.

We now consider the three-dimensional space Ω of the oriented line elements P situated on S. Let p denote the bearer point of P on S. Let, furthermore, P_t denote the line element obtained by moving P by arc length t tangentially along the geodesic through P. P_t depends continuously on P and t. For arbitrary fixed t, the transition from P to P_t is a one-to-one mapping of Ω into itself that is well known to leave the volume differential

$$dm = dod\varphi$$

invariant where do and $d\varphi$ denote the differentials of area and angle on S, respectively. Now, the function u constructed above is a bounded function of P, $u = u(P)$. For P fixed, $u(t) = u(P_t)$ is differentiable and satisfies the Riccati equation

$$\frac{du(P_t)}{dt} + u^2(P_t) + K(P_t) = 0 \qquad (K(P) = K(p)). \tag{12}$$

$u(P)$ is easily shown to be measurable (by a somewhat longer argument one can even prove its continuity). Equation (7) can be written

$$u(P) = \lim_{n=\infty} \frac{\dfrac{dY_n(s;\ P)}{ds}}{Y_n(s;\ P)}\Bigg|_{s=0}, \tag{13}$$

where $Y_n(s;\ P)$ is the solution of the Jacobi equation

$$Y''(s) + K(P_s)Y(s) = 0$$

that vanishes at P_n, i.e., at $s = n$, with the derivative one. It follows from the well-known continuity theorem for the solution of the initial value problem (which applies here as S is of class C''') that $Y_n(s;\ P)$ and its derivative with respect to s are continuous functions of P for s fixed. As $Y_n < 0$ for $s < n$ the fraction on the right in (13) depends continuously

on P. As a limit function of a sequence of continuous functions $u(P)$ must be measurable. Together with the boundedness this implies the summability of $u(P)$ over Ω in the sense of the measure m.

From (12) we obtain on integrating with respect to t

$$u(P_1) - u(P) + \int_0^1 K(P_t)dt = \int_0^1 u^2(P_t)dt \qquad (14)$$

for every P. Now, the invariance of the measure m under the mapping $P \to P_t$ implies the equality

$$\int_\Omega f(P_t)dm_P = \int_\Omega f(P)dm_P$$

for any summable function $f(P)$. By integration of (14) with respect to P and by using the last identity one obtains on the left-hand side

$$\int_\Omega [\int_0^1 K(P_t)dt]dm_P = \int_0^1 [\int_\Omega K(P_t)dm_P]dt = \int_\Omega K(P)dm$$
$$= 2\pi \int_S K(p)do.$$

The order of integration may be changed since $K(P_t)$ depends continuously on P and t. The right-hand side of (14) must be a summable function of P. The resulting equality

$$2\pi \int_S K(p)do = -\int_\Omega [\int_0^1 u^2(P_t)dt]dm_P$$

immediately proves the validity of the theorem. If the total curvature is zero the right-hand side of (14) must be zero for almost all P. For every such P, $u(P_t)$ must vanish at every t, $0 \leqq t \leqq 1$, because it depends continuously on t. From (12) one infers that $K(P) = 0$ for such a P. As a continuous function K must therefore vanish everywhere on S.

* The German original of this note had been dedicated and presented to C. Caratheodory on his seventieth birthday on September 13, 1943. Though accepted for subsequent publication it never appeared in print. Its loss by an air raid did not become known to the author until long after the end of the war.

[1] Morse, M., and Hedlund, G. A., "Manifolds Without Conjugate Points," *Trans. Am. Math. Soc.*, **51**, 362–386 (1942).

[2] Hopf, E., "Statistik der Loesungen geodaetischer Probleme vom unstabilen Typus," *Math. Annalen*, **117**, 590–608, in particular p. 608 (1940/41).

THE COMPLETENESS OF THE IRREDUCIBLE UNITARY REPRESENTATIONS OF A LOCALLY COMPACT GROUP

BY F. I. MAUTNER

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated by Oswald Veblen, December 16, 1947

The purpose of this note is to generalize some of the known results on unitary representations of compact and locally compact commutative groups to arbitrary separable locally compact groups. It is clear from such examples as the infinite cyclic group or the additive group of real numbers that unitary representations of non-compact groups in general do not decompose in a discrete manner. One needs a notion of generalized direct sum ("direct integral") of Hilbert spaces such as[1] von Neumann's[2] axiomatic characterization of generalized direct sums which he proves to be equivalent to the following:

Let $s(t)$ be a non-decreasing, right-semicontinuous, bounded, real-valued function of a real variable t and suppose that for each t a Hilbert space $\mathfrak{H}_t$ of finite or countably infinite dimension $k(t)$ is given such that $k(t)$ is s-measurable. The Hilbert space $\mathfrak{H}$ consisting of all s-measurable complex valued functions $x_n(t)$ $(n = 1, 2, \ldots, k(t))$ satisfying

$$\sum_n \int |x_n(t)|^2 ds(t) < \infty$$

is von Neumann's generalized direct sum or direct integral of the spaces $\mathfrak{H}_t$ with respect to the weight function $s(t)$: $\mathfrak{H} = \int_{\bullet} \mathfrak{H}_t$. Thus to every $x \in \mathfrak{H}$ corresponds a function $x(t)$ with values in the various $\mathfrak{H}_t$: $x \sim x(t)$. We shall call $x(t)$ the *component* of x in $\mathfrak{H}_t$. If x is given then $x(t)$ is determined for[3] almost all t and conversely. A bounded operator A in $\mathfrak{H}$ can now be said to decompose if $Ax \sim A(t)x(t)$ where $A(t)$ is for each t a bounded operator in $\mathfrak{H}_t$. (Compare reference 2.) Let F be any family of bounded operators A in $\mathfrak{H}$ and suppose each $A \in F$ decomposes into $A(t)$ under the given decomposition of $\mathfrak{H}$. Then we shall denote the family of operators $A(t)$ by $F(t)$.

THEOREM 1. *Let F be any self-adjoint family of bounded operators in a separable Hilbert space $\mathfrak{H}$. Then there exists a generalized direct sum $\mathfrak{H} = \int_{\bullet} \mathfrak{H}_t$ under which every $A \in F$ decomposes such that almost[3] all $\mathfrak{H}_t$ are irreducible under $F(t)$, i.e., have no proper closed invariant linear subspaces.*

The proof consists in showing that the various decompositions of $\mathfrak{H}$ into *irreducible* spaces $\mathfrak{H}_t$ correspond exactly to the *maximal* commutative subrings of the ring F' of those operators which commute with every $A \in F$.

THEOREM 2. *Let U be a unitary representation of a group G in a separable*

Hilbert space $\mathfrak{H}$. *Then* U *is generalized direct sum of irreducible unitary representations* U_t. *Suppose further that* G *is locally compact; then if* U *is continuous in the strong topology for operators*[6] *so are almost*[3] *all representations* U_t.

If $g \in G$ and $U(g)$ represent g then under the decomposition $\mathfrak{H} = \int_{\bullet} \mathfrak{H}_t$ the operator $U(g)$ is decomposed into operators $U(g, t)$ acting in $\mathfrak{H}_t$. According to reference 2 a change of $U(g, t)$ on a t-set of s-measure zero does not affect $U(g)$, thus to $U(g)$ corresponds for every g a *class* of (operator-valued) functions $U(g, t)$ and the proof consists in showing first that a representative can be chosen from each class in such a manner that the mapping $U(g) \rightarrow U(g, t)$ preserves the algebraic operations on operators, for all t outside a certain s-null-set which does not depend on g. Application of Fubini's theorem shows next that $U(g, t)$ is measurable (and hence continuous) in g for[3] almost all t.

The above implies that the completeness relation which was established by H. Weyl[4] for the space of square integrable functions on any coset-space of a compact group generalizes to *any* (separable) unitary representation space $\mathfrak{H}$ of an arbitrary locally compact group as follows:

If $\mathfrak{H} = \int_{\bullet} \mathfrak{H}_t$ is a decomposition of $\mathfrak{H}$ into irreducible representation spaces $\mathfrak{H}_t$ and if $y(t)$, $z(t)$ are the components of y, z in the space $\mathfrak{H}_t$ then

$$(y, z) = \int_{-\infty}^{\infty} (y(t), z(t)) ds(t), \tag{1}$$

where (y, z) denotes the inner product in $\mathfrak{H}$ and $(y(t), z(t))$ the inner product in $\mathfrak{H}_t$.

In particular if $\mathfrak{H}$ is the space $L_2(G)$ of square integrable functions $y(g)$, $z(g)$, ... on a separable locally compact group G and if $y(t)$, $z(t)$, are the components of $y(g)$, $z(g)$ in the (not necessarily finite dimensional) irreducible spaces $\mathfrak{H}_t$ then

$$\int_G y(g)\overline{z(g)} dg = \int_{-\infty}^{\infty} (y(t), z(t)) ds(t).$$

Here dg refers to left- or right-invariant Haar measure on G according as left- or right-translations in $L_2(G)$ are considered. In the sense of this Parseval relation those irreducible unitary representations of G which occur in a decomposition of the regular representation are complete in the space $L_2(G)$.

Another consequence of Theorem 2 is that every Haar-measurable positive definite function $f(g)$ on a separable locally compact group G can be expressed for all g outside a certain set of Haar-measure zero in the form

$$f(g) = \int_{-\infty}^{\infty} f_t(g) ds(t),$$

where the $f_t(g)$ are "elementary" positive definite functions on G in the sense of reference 5 and $s(t)$ a function of the type considered above,

depending on $f(g)$. This follows at once if one uses the correspondence established by Gelfand and Raikov[5] between the unitary representations of G and the positive definite functions on G, together with relation (1) in the form

$$(U(g)y, y) = \int_{-\infty}^{\infty}(U(g, t)y(t), y(t))ds(t).$$

A detailed account of the above results will be published elsewhere.

[1] I am very much indebted to Professor John von Neumann for the privilege of having been able to read his manuscript on "Generalized Direct Sums" before its publication.

[2] von Neumann, John, "On Rings of Operators. Reduction Theory." To appear soon in *Ann. Math.*

[3] I.e., for all t outside a certain set whose measure with respect to the weight function $s(t)$ is zero.

[4] Weyl, Hermann, "Harmonics on Homogeneous Manifolds," *Ibid.*, **35**, 486–499 (1934).

[5] Gelfand, I., and Raikov, D., "Irreducible Unitary Representations of Locally Bicompact Groups," *Compt. rend. acad. sci. U.R.S.S.*, **42**, 199–201 (1944); *Rec. Math. (Math. Sbornik)*, **13** (55), 301–316 (1943).

[6] von Neumann, John, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren," *Math. Annalen*, **102**, 370–427 (1929).

ON LACUNARY TRIGONOMETRIC SERIES, II

BY R. SALEM AND A. ZYGMUND

MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND UNIVERSITY OF CHICAGO

Communicated by Marshall Stone, December 2, 1947

This note is a continuation of another one with which the reader is supposed to be acquainted.[1] Except when otherwise stated, notation here is the same as before.

(vii) *Propositions* (i)–(vi) *remain valid even if the integral character of the numbers n_k in the lacunary series*

$$\sum_{k=1}^{\infty}(a_k \cos n_k x + b_k \sin n_k x),\ n_{k+1}/n_k > q > 1 \qquad (1)$$

is not assumed.

We shall confine our attention to (i) and (iii) which are typical. In the proof of (i) the integral character of the n_k played no rôle, so that the result obviously holds in the general case, even if instead of the distribution function of S_N/C_N in $(0, 2\pi)$ we consider the distribution function in any finite and fixed interval. As to (iii), its extension to non-integral n_k remains valid for every bounded set E of positive measure. The main difficulty in the proof of the extension is that the functions $\cos \gamma x$ which occur in the proof

of (iii) need not be orthogonal over any interval, which prevents us from using Bessel's inequality. To avoid this difficulty we may arrange the argument as follows.

First of all, a moment's consideration shows that it is enough to prove the result in the case when E is an interval. For then the result follows for E consisting of a finite or denumerable set of non-overlapping intervals, that is to say for open sets. Since every measurable set E is contained in an open set of measure arbitrarily close to $|E|$, we get the result for measurable sets. Thus let $E = (a, b)$, and let $K_\eta(x)$ be the familiar trapezoidal function equal to 1 in $(a + \eta, b - \eta)$, vanishing outside (a, b) and linear elsewhere. The area of the trapezoid we denote by k_η. If $Z_N(y)$ is the set of the points x at which $S_N(x)/C_N \leq y$, we define the function $H_N(y) = H_{N,\eta}(y)$ as

$$k_\eta^{-1} \int_{Z_N(y)} K_\eta(x)dx.$$

Clearly, $H_N(y)$ is a distribution function. It is enough to prove that it tends to $G(y) = (2\pi)^{-1/2} \int_{-\infty}^{y} \exp.(-1/2t^2)\, dt$, for then, making $\eta \to 0$, we obtain that the distribution function of $S_N(x)/C_N$ for (a, b) also tends to $G(y)$. For simplicity, we assume that (1) is a cosine series.

The distribution function of $H_N(y)$ is (see the proof of (iii))

$$\int_{-\infty}^{+\infty} e^{i\lambda y} dH_N(y) = k_\eta^{-1} \int_a^b K_\eta(x) e^{i\lambda S_N(x)/A_N} dx =$$

$$e^{-1/2\lambda^2} k_\eta^{-1} \int_a^b e^{o(1)} K_\eta(x) \left\{ \prod_{k=1}^{N} \left(1 + \frac{i\lambda a_k}{A_N} \cos n_k x \right) \right\} e^{1/2\lambda^2 \xi_N} dx,$$

where

$$\xi_N(x) = (2A_N^2)^{-1} \sum_{k=1}^{N} a_k^2 \cos 2n_k x.$$

The set of the points x in (a, b) at which $|\xi_N(x)| \geq \delta > 0$, is of measure

$$\leq \delta^{-2} \int_a^b \xi_N^2 dx \leq 1/4 \delta^{-2} A_N^{-4} \left[\sum_{k=1}^{N} (b - a) a_k^4 + \right.$$

$$\left. \sum_{1 \leq k \neq l \leq N} 2a_k^2 a_l^2 |n_k - n_l|^{-1} \right]. \quad (2)$$

Since $|n_k - n_l|$ stays above a positive number for $k \neq l$, and is large if either k or l is large, the second sum in square brackets is obviously $o(A_N^4)$. Since the first sum is also $o(A_N^4)$, the whole expression (2) is o (1). Thus, as $N \to +\infty$ and $\lambda = O(1)$, the characteristic function of $H_N(y)$ is, with an error $o(1)$,

$$e^{-1/2\lambda^2} k_\eta^{-1} \int_a^b K_\eta(x) \prod_{k=1}^{N} \left(1 + \frac{i\lambda a_k}{A_N} \cos n_k x \right) dx = e^{-1/2\lambda^2} k_\eta^{-1} I_N,$$

and it is enough to prove that $I_N \to k_\eta$. To show this we need the following

two lemmas whose proofs are almost identical with the proofs of Lemmas 1 and 2

LEMMA 1′. *Let a be any non-negative number. The number of solutions of the double inequality $a \leqslant n_{k_1} + n_{k_2} + \ldots + n_{k_p} \leq a + 1 (k_1 > k_2 > \ldots > k_p)$ is less than C^p, where $C = C(q)$.*

LEMMA 2′. *The number of solutions of $a \leq n_{k_1} + \ldots + n_{k_p} - n_{k_{p+1}} \leq a + 1$ $(k_1 > \ldots > k_p > k_{p+1})$ is less than C^{p+1}, where $C = C(q)$.*

Let us now consider the formula

$$\prod_{k=1}^{N}\left(1 + \frac{i\lambda a_k}{A_N}\cos n_k x\right) = \sum_{\nu} \alpha_\nu \cos \nu x, \qquad (3)$$

and let us set $\epsilon_k = |i\lambda a_k / A_N|$. Hence α_ν and ϵ_k depend also on N and λ. The numbers ν that occur in (3) are the non-negative, and in general non-integral, numbers of the form $\pm n_{k_1} \pm n_{k_2} \pm \ldots \pm n_{k_p}$, with $k_1 > k_2 > \ldots > k_p$. Let β_m denote the sum of the numbers $|\alpha_\nu|$ for $m < \nu \leqslant m + 1$, and let $m = 0, 1, \ldots$. As in the proof of (iii), we are led to consider equations of the form

$$n_{h_1} + n_{h_2} + \ldots + n_{h_s} = \nu + n_{l_1} + n_{l_2} + \ldots + n_{l_t}, \qquad (4)$$

and we get for β_m the estimate $2^{-(p-1)}\Sigma \epsilon_{h_1} \ldots \epsilon_{h_s}\epsilon_{l_1} \ldots \epsilon_{l_t}$ where $p = s + t$, and the summation is extended over all the solutions of (4) with $m < \nu \leq m + 1$. Arguing as before, we find that the β_m tend uniformly to 0 as $N \to \infty$ and $\lambda = O(1)$. Also $\alpha_0 - 1 \to 0$. On the other hand,

$$h_\nu = \int_a^b K_\eta(t) \cos \nu t \, dt = O(\nu^{-2}),$$

so that $|h_\nu| \leq M/(1 + \nu^2)$ for $\nu \geqslant 0$. Hence

$$I_N = \int_a^b K_\eta(t) \prod_{k=1}^{N}\left(1 + \frac{i\lambda a_k}{A_N}\cos n_k t\right)dt = k_\eta + (\alpha_0 - 1)k_\eta + \sum_{\nu>0}\alpha_\nu h_\nu,$$

$$|I_N - k_\eta| \leqslant |\alpha_0 - 1\,|k_\eta + M \sum_{m=0}^{\infty} \frac{\beta_m}{1 + m^2},$$

and this shows that $I_N \to k_\eta$. This completes the proof of the theorem.

Let us revert to the case when the n_k in (1) are positive integers satisfying an inequality $n_{k+1}/n_k > q > 1$. Instead of the partial sums S_N of the series (1) we may consider linear means of the partial sums. If we want to display the terms of the series we can always write the means in the form

$$\sigma_m(x) = \sum_{k=1}^{\infty} \alpha_{mk}(a_k \cos n_k x + b_k \sin n_k x) \qquad (5)$$

where the index m tends to $+\infty$ (in a continuous or discontinuous way). We assume that all the series (5) converge almost everywhere, and this implies that

$$\sum_k \alpha_{mk}^2(a_k^2 + b_k^2) < +\infty \text{ for every } m. \tag{6}$$

(This is not a restrictive condition since it is well known that if a lacunary trigonometric series is summable by any linear method of summability in a set of positive measure, the series must converge almost everywhere.) About the numbers α_{mk} we make only one assumption which will be retained in all what follows, namely that

$$(T) \qquad \lim_{m \to \infty} \alpha_{mk} = 1 \text{ for } k = 1, 2, \ldots$$

We also set

$$C_m^* = \left\{ {}^1/_2 \sum_k \alpha_{mk}^2 c_k^2 \right\}^{1/2}. \tag{7}$$

(viii) *Suppose that* $C_m = \left\{ {}^1/_2(c_1^2 + \ldots + c_m^2) \right\}^{1/2} \to \infty$, *and that*

$$\operatorname*{Max}_{1 \le j < \infty} \left\{ |c_j \alpha_{mj}| / C_m^* \right\} \to 0 \text{ as } m \to \infty. \tag{8}$$

Then the distribution function of $\sigma_m(x)/C_m^*$ *over any set* E, $|E| > 0$, *tends to the Gaussian distribution with mean value zero and deviation 1.*

The proof closely follows that of (iii). For, first of all, under condition (T) the divergence of Σc_m^2 implies that $C_m^* \to \infty$ (and conversely). Arguing as in the proof of (iii) and confining ourselves for the sake of simplicity to purely cosine series (so that we can write a_m, A_m, A_m^* for c_m, C_m, C_m^*) we see that the distribution function of $\sigma_m(x)/A_m^*$ is

$$|E|^{-1} \int_E e^{i\lambda \sigma_m(x)/A_m^*} dx = o(1)$$

$$+ e^{-\lambda^2/2} |E|^{-1} \int_E \prod_{k=1}^{\infty} \left(1 + i \frac{\lambda a_k \alpha_{mk}}{A_m^*} \cos n_k x \right) dx = o(1) + e^{-\lambda^2/2},$$

and the proof concludes as before.

Condition (8), though natural, is not quite simple, and we shall consider two special cases.

(*a*) *Suppose that for each* m *the numbers* α_{mk} *are non-negative and decrease as* k *increases. Then condition* (8) *is satisfied if* $c_j/C_j \to 0$.

For

$$c_j^2 \alpha_{mj}^2 / C_m^{*2} \le 2 c_j^2 \alpha_{mj}^2 / \sum_{k=1}^{j} \alpha^2{}_{mk} c_k^2 \le 2 c_j^2 / C_j^2.$$

Thus the left-hand side here is small for j large, uniformly in m. If j is not large, the left-hand side is small as $m \to \infty$ since the numerator $c_j^2 \alpha^2{}_{mj}$ is $\le c_j^2 \alpha_{mo}^2$ and the denominator C_m^{*2} tends to $\to \infty$.

The assumptions of (*a*) are satisfied by the methods of Cesaro of positive order and by the method of Abel. Considering, for example, the latter we conclude that

(ix) *If* $c_m \to \infty$, $c_m/C_m \to 0$, *and if we set* $C(r) = \left\{\frac{1}{2}\Sigma(a_k^2 + b_k^2)r^{2n_k}\right\}^{1/2}$,

$$f_r(x) = \sum_{k=1}^{\infty} (a_k \cos n_k x + b_k \sin n_k x) r^{n_k}$$

then, as $r \to 1$, *the distribution function of* $f_r(x)/C(r)$ *over any set* E, $|E| > 0$, *approaches the Gaussian distribution with mean value* 0 *and deviation* 1.

For $n_{k+1}/n_k \to \infty$, and $E = (0, 2\pi)$, see a paper by Kac.[2]

(*b*) *Suppose that the numbers* α_{mk} *are uniformly bounded in* m *and* k. *Then condition* (8) *is satisfied for* $c_j = O(1)$, $C_j \to \infty$.

For then the numerators in (8) are bounded and the denominators tend to $+\infty$, on account of condition (T). An example is provided by the Lebesgue method of summability which defines the sum of (1) as the limit of

$$\sum_{k=1}^{\infty} (a_k \cos n_k x + b_k \sin n_k x) \left(\frac{\sin n_k h}{n_k h}\right)$$

for $h(=1/m) \to +0$. Thus, if $c_k = O(1)$, $C_k \to \infty$, and if $F(x)$ is the sum of the series (1) integrated term by term (in view of the assumptions the series defining F converges absolutely and uniformly) then the distribution function over E of the differential ratio $\{F(x+h) - F(x-h)\}/2h$, divided by $\left\{\frac{1}{2}\Sigma c_k^2 \left(\frac{\sin n_k h}{n_k h}\right)^2\right\}^{1/2}$ approaches the standard Gaussian distribution. The same result holds for the ratio $\{F(x+2h) - F(x)\}/2h$ (and is a corollary of the preceding result if E is an interval). A simple illustration is provided by the Hardy-Weierstrass non-differentiable function $F(x) = \Sigma a^{-n} \cos a^n x$, $a > 1$.

Obviously a result parallel to (viii) holds for the series (1) of the class L^2. Since such series converge almost everywhere, we consider the distribution functions of the differences

$$\Delta_m(x) = f(x) - \sigma_m(x) = \sum_{k=1}^{\infty} \beta_{mk}(a_k \cos n_k x + b_k \sin n_k x)$$

where for each k the numbers $\beta_{mk} = 1 - \alpha_{mk}$ approach 0 as $m \to \infty$, on account of condition (T). Let us set

$$D_m = \left(\frac{1}{2}\sum_{k=m}^{\infty} c_k^2\right)^{1/2}, \qquad D_m^* = \left(\frac{1}{2}\sum_{k=1}^{\infty} \beta_{mk}^2 c_k^2\right)^{1/2}.$$

If we want each of the expressions $\sigma_m(x)$, $\Delta_m(x)$ to have meaning (almost everywhere in x) we must assume that the numbers D_m^* are finite, which is a consequence of the convergence of Σc_k^2 only if the elements of each row of the matrix $\{\alpha_{mk}\}$ are bounded. An obvious analogue of (viii) is

(x) *If the series* (1) *is of the class* L^2, *the* D_m^* *are finite, and*

$$\operatorname*{Max.}_{1 \leq j < \infty} \{|c_j \beta_{mj}|/D_m^*\} \to 0 \text{ as } m \to \infty, \qquad (9)$$

then the distribution function of $\Delta_m(x)/D_m^*$ *over any set* E, $|E| > 0$, *tends to the standard Gaussian distribution.*

As in the case (*a*) above, condition (9) is a consequence of $c_m/D_m \to 0$ if the numbers α_{mk} are non-negative and decrease as k increases. In particular if $c_m/D_m \to 0$, the distribution function of $\{f(x) - f_r(x)\}/\{{}^1/_2 \Sigma c_k^2 (1 - r^{n_k})^2\}^{1/2}$ tends to the standard Gaussian distribution.

We shall now consider besides (1) another lacunary series

$$\Sigma(a_k' \cos m_k x + b_k' \sin m_k x), \qquad m_{k+1}/m_k > q > 1, \qquad (10)$$

whose partial sums we shall denote by $S_M'(x)$.

(xi) *Suppose that* $\Sigma c_k^2 = \infty$, $\Sigma c_k'^2 = \infty$, *that* $c_k = o(C_k)$, $c_k' = o(C_k')$ *and that the combined sequence of the* n_k *and* m_k *is still lacunary. Then on every set* $E \subset (0, 2\pi)$, $|E| > 0$, *the expressions* $S_N(x)/C_N$, *and* $S_M'(x)/C_M'$ *are asymptotically independent as* $N, M \to \infty$.

To define the notion of asymptotic independence of S_N/C_N and S_M'/C_M' over E, let $Z_N(\alpha, \beta)$ denote the set of points $x \epsilon E$ such that $\alpha \leqslant S_N/C_N \leqslant \beta$, and let $Z_M'(\alpha', \beta')$ have a similar meaning. Let $Z_{NM}(\alpha, \beta;\ \alpha', \beta')$ be the set of points of E where the inequalities for S_N/C_N and S_M'/C_M' are satisfied simultaneously. The assertion of (xi) is that

$$\lim_{N, M \to \infty} \{|Z_{NM}(\alpha, \beta;\ \alpha', \beta')|/|E|\} = \lim_{N \to \infty} \{|Z_N(\alpha, \beta)|/|E|\} \lim_{M \to \infty} \{|Z_M'(\alpha', \beta')|/|E|\}$$

for all $\alpha, \beta, \alpha', \beta'$, the existence of the limits on the right being assured by (iii).

It is enough to sketch the proof of (xi) since it is analogous to that of (vi). It is enough to assume that $\alpha = \alpha' = -\infty$. Let us consider the joint distribution function $F_{NM}(\xi, \eta)$ of S_N/C_N and S_M'/C_M' over E, that is $Z_{NM}(-\infty, \xi;\ -\infty, \eta)/|E|$. The characteristic function of F_{NM} is

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i(\lambda\xi + \mu\eta)} dF_{NM}(\xi, \eta) = |E|^{-1} \int_E e^{i\{\lambda S_N(x)/C_N + \mu S_M'(x)/C_M'\}}\, dx.$$

Since we assume that the combined sequence of the n_k and m_k is lacunary the argument used in the proof of (vi) is applicable without change, and the last integral tends to exp. $\{-{}^1/_2(\lambda^2 + \mu^2)\}$. Thus

$$F_{NM}(\xi, \eta) \to (2\pi)^{-1} \int_{-\infty}^{\xi} \int_{-\infty}^{\eta} e^{-{}^1/_2(\lambda^2 + \mu^2)}\, d\lambda d\mu = (2\pi)^{-{}^1/_2} \int_{-\infty}^{\xi} e^{-{}^1/_2\lambda^2}\, d\lambda \cdot (2\pi)^{-{}^1/_2} \int_{-\infty}^{\eta} e^{-{}^1/_2\mu^2} d\mu,$$

and the factors on the right are asymptotically equal to $|Z_N(-\infty, \xi)|/|E|$ and $|Z_M'(-\infty, \eta)|/|E|$.

Various extensions of (xi) are possible. We might compare the remainders of the series (1) and (10) if the series are of the class L^2; or the partial

sums of one series with the remainders of the other; we might consider linear means of the series; or, finally, we might extend the results to the case of several series, provided the sequence of all exponents is still lacunary. The proofs would not require new ideas.

We shall now consider the case when the n_k in (1) are any set of linearly independent numbers, and again we confine our attention to an extension of (iii). Thus the condition of lacunarity is dropped and the n_k need not increase with k. By the distribution function F_N of $S_N(x)/C_N$ we mean $\lim_{T\to\infty}(2T)^{-1}|Z_N(y; -T, +T)|$, where $Z_N(y; -T, T)$ denotes the set of points x of the interval $(-T, T)$ at which $S_N/C_N \leq y$. The function $F_N(y)$ exists except for an at most denumerable set of y's, since this is true for the general almost periodic function in the sense of Bohr.[3]

(xii) *If the n_k are linearly independent, then the distribution function $F_N(y)$ of S_N/C_N tends to the standard Gaussian distribution provided $C_N \to \infty$, $c_N = o(C_N)$.*[4]

For let us again assume that we deal with a cosine series. Then

$$\int_{-\infty}^{+\infty} e^{i\lambda y}dF_N(y) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{i\lambda S_N(x)/A_N}\,dx =$$

$$e^{-\lambda^2/2}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\prod_{k=1}^{N}\left(1+\frac{i\lambda a_k}{A_N}\cos n_k x\right)dx + o_N(1) = e^{-\lambda^2/2} + o_N(1),$$

where $o_N(1)$ stands for a quantity small for N large, uniformly in T. The argument here is essentially the same as always. In particular, if $\xi_N(x)$ has the same meaning as before, and if $Z_N^*(\delta, -T, +T)$ denotes the set of x in $(-T, T)$ such that $|\xi_N(x)| \geq \delta$, then

$$\limsup_{T\to\infty}\frac{|Z_N^*(\delta,-T,T)|}{2T} = \limsup_{T\to\infty}\frac{1}{2\delta^2 T}\int_{-T}^{T}\xi_N^2 dx = \frac{1}{8\delta^2 A_N^4}\sum_{k=1}^{N} a_k^4,$$

and the last expression is small with $1/N$.

We conclude the note with a discussion of the behavior, for $N \to \infty$, of the sum

$$S_N(x) = a_1 f(n_1 x) + a_2 f(n_2 x) + \ldots + a_N f(n_N x)$$

where $n_1 < n_2 < \ldots$ are positive integers, the a_k are given, and $f(x)$ is a trigonometric polynomial with mean value zero. For the sake of brevity we assume that $f(x)$ is purely cosine,

$$f(x) = c_1 \cos lx + \ldots + c_m \cos mx,$$

and write

$$(2\pi)^{-1}\int_0^{2\pi} f^2 dx = {}^1/_2(c_1^2 + \ldots + c_m^2) = {}^1/_2 C, A_N = \left\{{}^1/_2 C(a_1^2 + \ldots + a_N^2)\right\}^{1/2}.$$

Thus $f^2(x) = {}^1/_2C + g(x)$, where $g(x)$ has mean value zero.

(xiii) *If* $A_N \to \infty$, $a_N/A_N \to 0$, *and*

$$n_{k+1}/n_k \geqslant g > m/l, \tag{11}$$

then the distribution of $S_N(x)/A_N$ *over any set* E, $|E| > 0$, *tends to the standard Gaussian distribution.*

We have to show that $|E|^{-1} \int_E \exp.\ (i\lambda S_N/A_N)dx \to \exp.\ (-{}^1/_2\lambda^2)$ as $N \to \infty$. Setting, as before,

$$\sum_1^N \frac{a_k^2}{A_N^2} f^2(n_kx) = 1 + \sum_1^N \frac{a_k^2}{A_N^2} g(n_kx) = 1 + \xi_N(x),$$

we must show that the set of points $x\epsilon(0, 2\pi)$ where $|S_N/A_N| \geqslant \delta > 0$ is of measure tending to 0 as $N \to \infty$, and this will follow if we prove that $\int_0^{2\pi} \xi_N^2 dx \to 0$. To compute this integral one has to consider the integrals of the type $\int_0^{2\pi} g^2(n_kx)dx = \int_0^{2\pi} g^2(x) = G$, and of the type $I_{h,k} = \int_0^{2\pi} g(n_hx)g(n_kx)dx$ for $h < k$. Since $g(x)$ is a polynomial of order $2m$ and mean value zero, $I_{h,k} = 0$ if $n_k > 2mn_h$. Therefore, for each h the number of the non-vanishing $I_{h,k}$ does not exceed an integer $r = r(m, q)$. Since $|I_{h,k}| \leqslant G$, one gets

$$\int_0^{2\pi} \xi_N^2 dx \leqslant BG(a_1^4 + \ldots + a_N^4)A_N^{-4} \to 0,$$

where $B = B(m, q)$. Arguing as in the proof of (iii), it remains to show that in the formula

$$\prod_1^N (1 + i\lambda a_k A_N^{-1} f(n_kx)) = \alpha_0^N + \Sigma \alpha_\nu^N \cos \nu x$$

we have $\alpha_0^N \to 1$, and $\alpha_\nu^N \to 0$ for $\nu > 0$.

LEMMA 1″. *The number of solutions of* $m_1n_{k_1} + m_2n_{k_2} + \ldots + m_pn_{k_p} = A$ *when* $l \leqslant m_i \leqslant m$, is $< D^p$, *where* $D = D(l, m, q)$.

LEMMA 2″. *The number of solutions of* $m_hn_h - m_kn_k = A$, *with* m_h *and* m_k *between* l *and* m, *is* $< E = E(l, m, q)$.

The proof of Lemma 1″ requires no explanation. In the proof of Lemma 2″, writing $A > ln_h - mn_k > ln_h(1 - m/lq)$ we use the fact that $q > m/l$. The proof of (xiii) is now completed by observing that if $\epsilon_k = |i\lambda a_k/A_N|$ the contribution in α_ν^N of all solutions of $m_1n_{h_1} + \ldots + m_sn_{h_s} = \nu + m_1'n_{j_1} + \ldots + m_t'n_{j_t}$ is at most

$$2^{-(p-1)} \sum \epsilon_{k_1} \ldots \epsilon_{k_f} \epsilon_{j_1} \ldots \epsilon_{j_t} |c_{m_1} \ldots c_{m_s} c_{m_1'} \ldots c_{m_t'}| \leqslant 2^{-(p-1)} (\text{Max. } |c_i|)^p \sum \epsilon_{h_1} \ldots \epsilon_{h_s} \epsilon_{j_1} \ldots \epsilon_{j_t}$$

where $p = s + t$. From this point the proof proceeds as before.

Remark. That condition (11) cannot be relaxed is seen on the example $f(x) = \cos x + \cos mx$, $n_k = m^k - 1$. For then

$$\cos mn_k x + \cos n_{k+1}(x) = \cos\left(m^k - \frac{m+1}{2}\right) x \cos\frac{1}{2}(m-1)x$$

and combining the second term of $f(n_k x)$ with the first term of $f(n_{k+1}x)$ we see that the Gaussian distribution is altered by the factor $\cos {}^1/_2(m-1)x$. This remark is essentially due to Erdös.

[1] These PROCEEDINGS, **33**, 333–338 (1947).

[2] *Amer. J. Math.*, 1939.

[3] Wintner, A., *Am. J. Math.*, **54**, 339–345 (1932); see also Jessen, B., *J. London Math. Soc.*, **8**, 247–250 (1933).

[4] Compare Kac and Steinhaus, *Studia Mathematica*, **4**, 11 (1938).

APPLICATIONS OF CYCLOTOMY TO THE THEORY OF NON-HOMOGENEOUS EQUATIONS IN A FINITE FIELD

BY H. S. VANDIVER

DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF TEXAS

Communicated December 22, 1947

In another paper[1] we obtained the following result:

"If p is an odd prime belonging to the exponent t modulo l and such that $p^t = 1 + cl$ with c and l prime to each other and l odd. Write $\theta = \zeta\beta$, where ζ is a primitive l-th root of unity and β is a primitive c-th root of unity. Let g be a primitive root of a finite field $F(p^t)$ of order p^t, and denote by $[v, r]$ the number of distinct solutions g^α, g^γ of

$$1 + g^{v+c\alpha} = g^{r+\gamma l}, \tag{1}$$

α in the set $0, 1, \ldots, l-1$ and γ in the set $0, 1, \ldots, c-1$. Associated with this number (v, r) is the cyclotomic number

$$\psi_{a,b}(\theta) = \sum_{i,j} [i, j]\beta^{bi}\zeta^{aj}, \tag{2}$$

with $a \not\equiv 0 \pmod{l}$, $b \not\equiv 0 \pmod{c}$, i ranging over the set $0, l, \ldots, c-1$ and j ranging over the set $0, 1, \ldots, l-1$. Said function has the property

$$\psi_{a,b}(\theta)\psi_{a,b}(\theta^{-1}) = p^t. \tag{3}$$"

In the statement of this result in the former paper we assumed that l was prime. This restriction was not necessary and was not employed in the proof of the theorem. We also stated in connection with this theorem that certain quadratic relations governing the solution of (1) could be developed with the use of (2) and (3). In the present paper we shall derive said quadratic relations and also consider their connection with some results in another paper.[2]

In view of (3) we have

$$\sum_{i',j'} [i', j']\beta^{bi'}\zeta^{aj'}\sum_{i,j} [i, j]\beta^{-bi}\zeta^{-aj} = p^t,$$

where i, i', range independently over $0, 1, \ldots, c-1$ and j, j' independently over $0, 1, \ldots, l-1$, whence

$$\sum_{i',j',i,j} [i', j'][i, j]\beta^{b(i'-i)}\zeta^{a(j'-j)} = p^t. \tag{4}$$

Set $i' - i = h$ and $j' - j = k$. Then (4) becomes

$$\sum_{i,j,h,k} [i + h, j + k][i, j]\beta^{bh}\zeta^{ak},$$

h ranging independently over $0, 1, \ldots, c-1$, and k independently over $0, 1, \ldots, l-1$.

Set

$$A_k = \sum_{i,j,h} [i + h, j + k][i, j]\beta^{bh}.$$

Then (4) gives

$$A_0 + \zeta^a A_1 + \zeta^{2a} A_2 + \ldots + \zeta^{(l-1)a} A_{l-1} = p^t. \tag{5}$$

We may take $a = 1, 2, \ldots, l-1$ and obtain $l-1$ equations. Also we have

$$A_0 + A_1 + A_2 + \ldots + A_{l-1} = 1, \tag{6}$$

since it is the product of

$$\sum_{i,j} [i, j]\beta^{bi} \text{ and } \sum_{i',j'} [i', j']\beta^{-bi'}$$

and each of these is (-1). Making an array of the l equations obtained from (5) and (6) and eliminating β therefrom by the use of a method of a previous paper,[3] gives

$$lA_0 = (l-1)p^t + 1,$$

or

$$A_0 = cl - c + 1 = p^t - c \tag{6a}$$

and

$$\begin{aligned} A_s &= -c, \\ s &= 1, 2, \ldots, l-1. \end{aligned} \tag{6b}$$

Set, if i and j range as before,

$$B_{h,k} = \sum [i + h, j + k][i, j]. \tag{7}$$

We shall now show that

$$B_{0,k} + B_{1,k} + B_{2,k} + \ldots + B_{c-1,k} = c^2l - 2c + 1 \qquad (8)$$

for $k = 0$. By the definition (7) the left-hand member of this equals

$$\sum_{i,j,h} [i + h, j][i, j] = \sum_{i,j} ([i, j][i, j] + [i + 1, j][i, j] + \ldots + [i + c - 1, j][i, j]). \qquad (9)$$

We have, by keeping v fixed in (1) and counting the number of possible values of γ,

$$\sum_{r=0}^{l-1} [v, r] = l \text{ if } v \neq \frac{c}{2}, \qquad (10)$$

and

$$\sum_{r=0}^{l-1} \left[\frac{c}{2}, r\right] = l - 1. \qquad (11)$$

Similarly

$$\sum_{v=0}^{c-1} [v, r] = c \text{ if } r \neq 0, \qquad (12)$$

and

$$\sum_{v=0}^{c-1} [v, 0] = c - 1. \qquad (13)$$

The expressions $i, i + 1, i + 2, \ldots, i + c - 1$ are congruent in some order to $0, 1, 2, \ldots, c - 1$ modulo c; and hence by (12) we have for a fixed $j \neq 0$ the sum of the corresponding terms of (9) equals $\sum_{i,j} c[i, j]$; while for $j = 0$ the sum of the corresponding terms of (9) is by (13) equal to $\sum_i (c - 1)\,[i, 0]$; whence (9) equals, $(j \neq 0)$,

$$\sum_{i,j} c[i, j] + \sum_i (c - 1)[i, 0] = \qquad (14)$$

$c^2(l - 1) + (c - 1)^2$ or $c^2l - 2c + 1$ by again applying (12) and (13). But this is the conclusion stated in (8).

For $k \neq 0$, we have

$$\begin{aligned}
&\sum_{h,i,j} [i + h, j + k][i, j] \\
&= \sum_{i,j} [i, j]([i, j + k] + [i + 1, j + k] + \ldots + [i + c - 1, j + k]) \\
&= \sum_{i,\, j \neq l-k} [i, j]c + \sum_i (c - 1)[i, l - k] \\
&= c \sum_{j \neq l-k} ([0, j] + [1, j] + \ldots + [c - 1, j]) + \sum_i (c - 1)[i, l - k] \\
&= c(c - 1) + c^2(l - 2) + c(c - 1) = c^2l - 2c.
\end{aligned}$$

Or for $k \neq 0$,

$$\sum_{h,i,j} [i + h, j + k][i, j] = c^2l - 2c. \qquad (15)$$

Employing (6a) and (14) we may write

$$\begin{aligned} &B_{00} + B_{10} + \ldots + B_{c-1,0} = c^2(l-1) + (c-1)^2, \\ &B_{00} + \beta B_{10} + \ldots + \beta^{c-1} B_{c-1,0} = p^t - c, \\ &\qquad\qquad \ldots\ldots\ldots \\ &\qquad\qquad \ldots\ldots\ldots \\ &B_{00} + \beta^{c-1} B_{10} + \ldots + \beta^{(c-1)(c-1)} B_{c-1,0} = p^t - c. \end{aligned}$$

Also for $k \not\equiv 0$, we have using (6b) and (15),

$$\begin{aligned} &B_{0k} + B_{1k} + \ldots + B_{c-1,k} = c^2 l - 2c \\ &B_{0k} + \beta B_{1k} + \ldots + \beta^{c-1} B_{c-1,k} = -c \\ &\qquad\qquad \ldots\ldots\ldots \\ &\qquad\qquad \ldots\ldots\ldots \\ &B_{0k} + \beta^{c-1} B_{1k} + \ldots + \beta^{(c-1)(c-1)} B_{c-1,k} = -c. \end{aligned}$$

Whence, employing the same method we used to obtain (6a) and (6b) from (5) and (6),

$$\left.\begin{aligned} B_{00} &= 2lc - l - c, \\ B_{h0} &= cl - l - 1 \text{ for } h \not\equiv 0 \pmod{c}, \\ B_{0k} &= cl - c - 1 \text{ for } k \not\equiv 0 \pmod{l}, \\ B_{hk} &= cl - 1 \text{ for } h \not\equiv 0 \pmod{c},\ k \not\equiv 0 \pmod{l}. \end{aligned}\right\} \tag{16}$$

Whence the

THEOREM I, *if p is an odd prime belonging to the exponent t modulo l and such that $p^t = 1 + lc$ with c and l prime to each other and l odd, and also g as a primitive root in a finite field $F(p^t)$. Denote by $[v, r]$ the number of distinct solutions g^α, g^γ of the equation*

$$1 + g^{v+c\alpha} = g^{r+l\gamma},$$

where α is in the set $0, 1, \ldots, l-1$ and γ in the set $0, 1, \ldots, c-1$. Also set

$$B_{h,k} = \sum_{i,j} [i + h, j + k][i, j].$$

Then the relations (16) follow.

In the relation (1) we took for the exponent c and l to have the property that $cl = p^t - 1$. However, we might have taken c_1 and l_1 in place of c and l in (1) where $(c_1, l_1) = 1$, and $c_1 l_1$ divides $p^t - 1$, then if $\{i, j\}$ denotes the number of distinct solutions of the equation

$$1 + g^{i+c_1\delta} = g^{j+l_1\epsilon},$$

then we may use in lieu of (2) the number

$$\psi_{a,\epsilon}(\theta_1) = \sum_{i,j} \{i, j\} \beta_1^{di} \zeta_1^{\epsilon j},$$

where β_1 is a primitive c_1-th root of unity and ζ_1 is a primitive l_1-th root of

unity. Also θ_1 is a primitive in $(c_1 l_1)$-th root of unity. This function also has the same property given in (3) and we may obtain relations analogous to (16) involving the quantities $\{i + h, j + k\}\{i, j\}$. As we were principally interested in the case where $cl = p^t - 1$ in (1), in view of certain applications, we did not carry out this more general investigation in detail.

In a former paper[2] we examined the number (i, j) of solutions g^r, g^s of

$$1 + g^{i+rl} = g^{j+sl},$$

where r and s are each in the range of $0, 1, \ldots, c - 1$ and obtained quadratic relations involving the quantities $(i, j)(j + h, j + k)$, given in Theorem I of that paper. The relations seem to be of somewhat different character from those of the present paper and we do not yet have a unified theory to cover the case where c and l are any integers in (1) of such a nature as to yield as special cases relations (16) of this paper as well as (23), (24), (25), (26) of the former paper.

[1] These PROCEEDINGS, **32**, 317 (1946).
[2] *Ibid.*, **33**, 236–242 (1947).
[3] *Ibid.*, **31**, 189 (1945).

A SPECIAL CLASS OF SOLUTIONS OF THE EQUATIONS OF THE GRAVITATIONAL FIELD ARISING FROM CERTAIN GAUGE-INVARIANT ACTION PRINCIPLES

BY H. A. BUCHDAHL

PHYSICS LABORATORY, UNIVERSITY OF TASMANIA

Communicated by Oswald Veblen, December 18, 1947

1. By direct consideration of the variation of the integrals

$$J_1 = \int G^2 \sqrt{-g}\, d\tau, \qquad J_2 = \int G^{\mu\nu} G_{\mu\nu} \sqrt{-g}\, d\tau$$

the author has shown[1] that in a V_4 the field equations arising out of the action principles $\delta J_1 = 0$, $\delta J_2 = 0$ are satisfied by any solution of the equations

$$G_{\mu\nu} = \alpha g_{\mu\nu}, \tag{1}$$

where α is an *arbitrary* constant, and $G_{\mu\nu}$ is the Ricci tensor. Let $\overline{B}_{\mu\nu\alpha}{}^{\beta}$ be the gauge-invariant curvature tensor of Weyl's theory;[2] and let $\overline{B}_{\mu\nu\alpha}{}^{\alpha} = \overline{G}_{\mu\nu}$, $g^{\mu\nu}\overline{G}_{\mu\nu} = \overline{G}$, so that in a region free from electromagnetic radiation $\overline{G}_{\mu\nu} = G_{\mu\nu}$, $\overline{G} = G$. Then by a slight extension of the method referred to above we may prove the following:

THEOREM. *In a region free from radiation* ($\kappa_\mu = 0$) *the field equations arising from a gauge-invariant action principle in which the Lagrangian is composed of the components of the contracted curvature tensor and the components of the metrical tensor are satisfied by any solution of the equations* $G_{\mu\nu} = \alpha g_{\mu\nu}$, *where* α *is an arbitrary constant.*

2. Defining the "absence of radiation" to mean

$$\kappa_\mu = 0, \qquad (\mu = 1, 2, \ldots, n) \tag{2}$$

we may prove the theorem somewhat more generally for a V_n. When (2) applies every action principle of the kind under consideration reduces to the form

$$\delta \int L(K_1, K_2, \ldots, K_m) \sqrt{-g}\, d\tau = 0 \tag{3}$$

where the Lagrangian is some function of the $K_s = G^{\mu_2}_{\mu_1} G^{\mu_3}_{\mu_2} G^{\mu_4}_{\mu_3} \ldots G^{\mu_1}_{\mu_s}$, $(K_1 = G)$, and m is a positive integer. Equation (3) is the degenerate form of the action principle $\delta \int \bar{L} \sqrt{-g} d\tau = 0$, where the Lagrangian $\bar{L}$ is composed of the $\bar{G}_{\mu\nu}$ and the $g^{\mu\nu}$. In a gauge transformation in which $g_{\mu\nu}$ becomes $\lambda^2 g_{\mu\nu}$, $\sqrt{-g}$ becomes $\lambda^{+n}\sqrt{-g}$ and $\bar{G}_{\mu\nu}$ remains unchanged, so that $\bar{L}$ must change into $\lambda^{-n}\bar{L}$ since $\bar{L}\sqrt{-g}$ is to be a gauge-invariant density. Putting $\kappa_\mu = 0$ after the transformation has been carried out this implies the condition

$$L(\mu K_1, \mu^2 K_2, \ldots, \mu^m K_m) \equiv \mu^{n/2} L(K_1, K_2, \ldots, K_m), \tag{4}$$

where we have set $\mu = \lambda^{-2}$.

Differentiating (4) with respect to μ and putting $\mu = 1$ after the differentiation, this gives

$$\sum_{s=1}^{m} sK_s \frac{\partial L}{\partial K_s} = {}^1/_2 nL. \tag{5}$$

Now

$$\delta L = \sum_{s=1}^{m} \frac{\partial L}{\partial K_s} \delta K_s; \tag{6}$$

and in view of the symmetry of K_s with respect to the $G^{\mu_{i+1}}_{\mu_i}$ we have

$$\delta K_s = sG^{\mu_2}_{\mu_1} G^{\mu_3}_{\mu_2} \ldots G^{\mu_s}_{\mu_{s-1}} \delta G^{\mu_1}_{\mu_s}.$$

If now $G_{\mu\nu}$ be given by (1) it follows immediately that

$$K_s = \text{const.} = n\alpha^s, \qquad \delta K_s = s\alpha^{s-1}\delta G = (sK_s/n\alpha)\delta G.$$

Hence, from (6),

$$\delta L = \frac{1}{n\alpha} \sum_{s=1}^{m} sK_s \frac{\partial L}{\partial K_s} \delta G = L\delta G/2\alpha,$$

by (5). Therefore

$$\delta\int L\sqrt{-g}d\tau = \int\left(\frac{1}{2\alpha}L\sqrt{-g}\delta G + L\delta\sqrt{-g}\right)d\tau$$
$$= (L/2\alpha)\int(\delta(\sqrt{-g}G) - G\delta\sqrt{-g} + 2\alpha\delta\sqrt{-g})d\tau,$$

or

$$\delta\int L\sqrt{-g}d\tau = (L/2\alpha)\delta\int(G - (n-2)\alpha)\sqrt{-g}d\tau. \quad (7)$$

Accordingly the condition that the right-hand side of (7) should vanish is always compatible with equation (1) whereby it was transformed into the condition expressed by the vanishing of the right-hand side of (7). For[3]

$$\delta\int(G - (n-2)\alpha)\sqrt{-g}d\tau = \int({}^1/_2 g^{\mu\nu}G - G^{\mu\nu} - {}^1/_2(n-2)\alpha g^{\mu\nu})\sqrt{-g}\delta g_{\mu\nu}d\tau = 0,$$

by (1). Hence $\delta\int L\sqrt{-g}d\tau$ vanishes, and the theorem of section 1 is proved.

In particular the result applies to the case where the action is taken to be the generalized volume $\int\sqrt{-\det|\overline{G}_{\mu\nu}|}d\tau$, so that

$$L = \sqrt{\det|G_{\mu\nu}|/g} \equiv (\sum a_{\nu_1\ldots\nu_m}K_1^{\nu_1}K_2^{\nu_2}\ldots K_m^{\nu_m})^{1/1}, \qquad (\sum_{r=1}^{m} r\nu_r = n),$$

where the a's are certain numerical coefficients. This clearly is of a form satisfying the condition (5).

[1] Buchdahl, H. A., *Proc. Edinburgh Math. Soc.* (in the press) (1948).

[2] Eddington, A. S., *The Mathematical Theory of Relativity*, Cambridge University Press, 1930, p. 204.

[3] Reference 2, p. 139.

PHYSICAL FAMILIES OF CURVES IN SPACE

BY EDWARD KASNER AND JOHN DE CICCO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, AND ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO

Read before the Academy, November 18, 1947

1. In preceding papers, we studied physical families of curves in the plane. We gave characteristic properties and discussed the transformation theory. In the present paper, we shall generalize this theory to space of three dimensions.

The important physical families of curves connected with an arbitrary (positional) field of force[1] are (1) trajectories, that is, the paths of motion of a particle; (2) brachistochrones, that is, the curves in a general (con-

servative) field of force along which the time of the constrained motion between any two points is least; (3) general catenaries, that is, the curves formed from the resulting equilibrium when an inextensible flexible string is suspended in the field; and finally (4) velocity systems, that is, a curve is a velocity curve corresponding to the speed v_0, provided a particle starting from any lineal element of the curve with that speed describes a trajectory osculating the curve.

2. In the plane, we have shown that these four systems of physical curves may all be obtained as special cases of the following general problem.[2] To find curves along which a constrained motion is possible such that the pressure P is proportional to the normal component N of the force. Thus a *system of curves* S_k for an arbitrary field of force is defined by the relation $P = kN$.

Each of these systems S_k is a three-parameter family and is defined by a differential equation of third order

$$(\psi - y'\phi)y''' = [(\psi_x + y'\psi_y) - y'(\phi_x + y'\phi_y)]y'' - \left[3\phi + \frac{(n-2)(\phi + y'\psi)}{1 + y'^2}\right] y''^2, \quad (1)$$

where $n = 2/(k+1)$. Of course (ϕ, ψ) are the rectangular components of the force vector and primes denote differentiation with respect to x.

3. Now we begin the investigation of the corresponding theory in space of three dimensions. A *system of curves* S_k in an arbitrary field of force with rectangular components (ϕ, ψ, χ) consists of curves along which a constrained motion is possible such that the osculating plane at each point contains the force vector and the pressure P is proportional to the normal component N of the force so that $P = kN$.

The tangential component T and the normal component N (along the principal normal) of the force vector are defined by the formulae

$$T = \frac{\phi + y'\psi + z'\chi}{(1 + y'^2 + z'^2)^{1/2}} \qquad N = \left[\frac{(\psi - y'\phi)^2 + (\chi - z'\phi)^2 + (y'\chi - z'\psi)^2}{1 + y'^2 + z'^2}\right]^{1/2}. \quad (2)$$

provided the force vector lies in the osculating plane. The radius of curvature r of a curve is

$$r = \frac{(1 + y'^2 + z'^2)^{3/2}}{[y''^2 + z''^2 + (y'z'' - y''z')^2]^{1/2}}. \quad (3)$$

If the particle is assumed to be of unit mass, the pressure P, along the principal normal to the path of the particle, is

$$P = \frac{v}{r^2} - N. \quad (4)$$

Along a curve of the system S_k, we must have $P = kN$. This condition $P = kN$ is equivalent to the two equations

$$P = \frac{v^2}{r} - N = kN, \quad v\frac{dv}{ds} = T, \tag{5}$$

where v denotes speed and s arc length. Eliminating the speed v from these two relations, we find the condition

$$N\frac{dr}{ds} = nT - r\frac{dN}{ds}, \tag{6}$$

where $n = \dfrac{2}{k+1}$.

Using the condition that the force vector lies in the osculating plane, we find from (2) and (3) that (6) reduces to the form

$$\frac{d}{dx}\left[\frac{1}{y''}(\psi - y'\phi)(1 + y'^2 + z'^2)\right] = n(\phi + y'\psi + z'\chi). \tag{7}$$

Simplifying this together with the condition that the force vector lies in the osculating plane, we find that *the differential equations defining any system S_k in an arbitrary field of force, are*

$$(\psi - y'\phi)y''' = [(\psi_x + y'\psi_y + z'\psi_z) - y'(\phi_x + y'\phi_y + z'\phi_z)]y'' - \left[3\phi + \frac{(n-2)(\phi + y'\psi + z'\chi)}{1 + y'^2 + z'^2}\right]y''^2, \tag{8}$$

$$z'' = \frac{\chi - z'\phi}{\psi - y'\phi}y''.$$

A system S_k in space consists of a five-parameter family of curves. Thus there are ∞^5 curves in a given S_k. The relation between k and n is

$$n = \frac{2}{k+1}. \tag{9}$$

It is found that *the four cases of physical interest are*

$k = 0$ or $n = 2$ gives S_0, the system of *trajectories*.
$k = -2$ or $n = -2$ gives S_{-2}, the system of *brachistochrones*.
$k = 1$ or $n = 1$ gives S_1, the system of *general catenaries*.
$k = \infty$ or $n = 0$ gives S_∞, the system of *velocity curves*.

4. We find that any system S_k of ∞^5 curves possesses the following *Geometric Properties I and II*.[3]

I. *The osculating planes of the ∞^3 curves passing through a given point form a pencil; that is, all the planes pass through a fixed direction.*

II. *The osculating spheres of the ∞^1 curves passing through a given point*

in a given direction form a pencil; their centers thus describe a straight line.

5. We also find that *the associated plane systems, determined by the given space system in the manner described below, have the five geometric properties characteristic of a planar system* S_k'.

Consider the given system S_k of ∞^5 space curves in connection with any plane p. Through any point of p there pass ∞^2 curves of the given system which are tangent to the plane. Project the differential elements of the third order belonging to these space curves orthogonally upon p, thus obtaining ∞^2 plane differential elements of the third order at the selected point. Applying this process to all points of p, we have a defined set of ∞^4 differential elements of the third order. These elements define a certain differential equation of the third order, and thus determine a system of ∞^3 integral curves. This we term the *associated system* in the plane p. The space system S_k has the property that every one of these plane systems associated with it is a planar system S_k'.

6. In the space system S_k, consider the curves which pass through a given point in the direction of the line of force. We find that all of these have zero curvature except one particular curve. *The ratio* ρ *of the curvature of this one particular curve of the system* S_k *to the curvature of the tangent line of force is*

$$\rho = \frac{1}{n+1} = \frac{k+1}{k+3}. \tag{10}$$

For the four cases of physical interest, this ratio ρ is as follows: (1) For the system S_0 of trajectories, we have $\rho = {}^1/_3$. This is the Theorem of Kasner on the ratio of departures of the rest trajectory and the line of force from their common tangent line. (2) For the system S_{-2} of brachistochrones, we find $\rho = -1$. (3) For the system S_1 of general catenaries, we obtain $\rho = {}^1/_2$. (4) For the system S_∞ of velocity curves, we get $\rho = 1$.

7. Finally we discuss the transformation theory of spacial systems S_k. We have established the following result.[4]

Except for the system S_0 *of trajectories, the only point transformations which convert every system of curves* S_k *into a system* S_k *are the similitudes which form a seven-parameter group in space. All the point transformations of space which send every system* S_0 *of trajectories into a system* S_0 *of trajectories form the fifteen-parameter collineation group.*

In particular, we find that the fundamental two-dimensional projective theory given by Appell (1889) can be extended to three dimensions.

Later we shall discuss the corresponding theory for generalized fields of force where the field of force depends not only on the position of the point but also on the direction at the point.

[1] Kasner, "Differential-Geometric Aspects of Dynamics," Princeton Colloquium Lectures, *Am. Math. Soc. Publications*, 1913 (1934) (1948).

[2] Kasner, "Physical Curves," these PROCEEDINGS, **33**, 246–251 (1947).

[3] Kasner, "The Trajectories of Dynamics," *Trans. Am. Math. Soc.*, **7**, 401–424 (1906). Also, "Dynamical Trajectories: The Motion of a Particle In An Arbitrary Field Of Force", *Trans. Am. Math. Soc.*, **8**, 135–158(1907).

[4] For the plane transformation theory, see Kasner and De Cicco, "Transformation Theory of Physical Curves," these PROCEEDINGS, **33**, 338–342 (1947).

---

*A PHYSIOLOGICAL BASIS FOR SOME SUPRESSOR MUTATIONS AND POSSIBLY FOR ONE GENE HETEROSIS**

BY STERLING EMERSON

KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIF.

Communicated by G. W. Beadle, December 13, 1947

A mutant strain of *Neurospora* which requires sulfonamides for growth at 35° frequently becomes altered so that it can grow at that temperature without sulfonamides.[1] In each instance analyzed, the developed ability to grow without sulfonamides has resulted from mutation of genes distinct from that responsible for the sulfonamide requirement. Each "reverted" strain has proved to be a heterocaryon, composed of two kinds of nuclei, both of which carry the gene (*sfo*) for sulfonamide requirement, but one also carries a new "supressor" gene. At least two independent supressor genes have been involved in such reversions. Homocaryotic strains have been isolated, some carrying both *sfo* and the supressor, others carrying the supressor alone. None of these isolates has the growth characteristics of wild type, but each shows some peculiar relationship to the *p*-aminobenzoic acid requirement, the nature of which is still obscure. In a few instances, artifically constructed heterocaryons between the sulfonamide-requiring strain and a strain carrying both the sulfonamide-requiring gene and a supressor have resulted in better growth on minimal medium than either strain is capable of by itself, thus reconstituting the situation observed in the original "reverted" heterocaryons.

As long as it was believed that sulfonamides were used as metabolites by the sulfonamide-requiring strain it was difficult to picture a possible physiological basis for the circumvention of the drug requirement by mutation of some entirely different gene. Recently Zalokar[2] has shown that the "sulfonamide-requiring" strain can grow in the absence of sulfonamides provided the available *p*-aminobenzoic acid is reduced to a particular concentration, growth being inhibited by both higher and lower

concentrations. It now appears that in the sulfonamide-requiring strain some reaction catalyzed by *p*-aminobenzoic acid is detrimental in that growth is prevented unless the effective amount of *p*-aminobenzoic acid is decreased, either by the competition of sulfonamides,[1] or by a reduction in the amount synthesized by the strain.[2] It follows then that any mutation

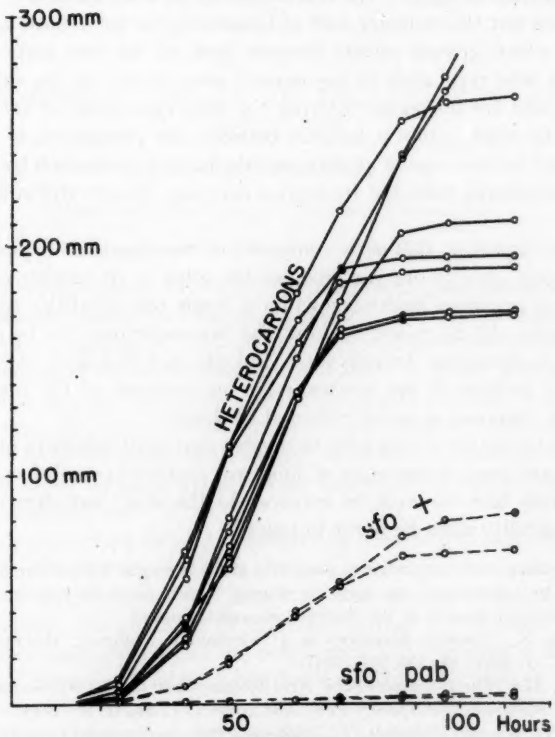

FIGURE 1

Growth curves of the sulfonamide-requiring mutant strain (*sfo*, +), the double mutant sulfonamide requiring, *p*-aminobenzoicless (*sfo*, *pab*), and of heterocaryons between them, on minimal medium at 35°. Changes in growth rates in the heterocaryons are presumably due to changing relative frequencies of the two types of nuclei at the growing tips.

which would have the effect of reducing the available *p*-aminobenzoic acid content might permit growth of the sulfonamide-requiring strain in the absence of sulfonamides, and this could happen even if the mutant gene was carried by only part of the nuclei in a heterocaryon.

To test this possibility, artificial heterocaryons were made between a double mutant strain carrying the sulfonamide-requiring gene (*sfo*) and a gene (*pab*) which prevents the synthesis of *p*-aminobenzoic acid, and a strain carrying *sfo* and the wild type allele (+) of *pab*. Neither of these strains *sfo*, *pab* or *sfo*, + grows appreciably on minimal medium at 35°, but as indicated in figure 1 the heterocaryons between them do.

These are not the ordinary sort of heterocaryons (cf. Beadle and Coonradt[3]) in which growth results because each of the two sorts of nuclei carries the wild type allele of the mutant gene carried by the other, since in these there are no nuclei carrying the wild type allele of *sfo*. In this case growth results from a balance between the production of *p*-aminobenzoic acid by one type of nucleus and the lack of production by the other to give an amount tolerated by strains carrying *sfo*, yet still sufficient for growth.

A heterocaryon of this sort, composed of two kinds of haploid nuclei, both carrying *sfo*, and one carrying *pab* the other +, is roughly equivalent to a diploid organism heterozygous for a single pair of alleles *pab*/+, the homocaryotic strains mixed to form the heterocaryons can be compared to the corresponding homozygous diploids *pab*/*pab* and +/+. The augmented growth of the heterocaryon reminds one of the instances of single gene heterosis in maize reported by Jones.[4]

It is not the intent of this note to suggest that most instances of one gene heterosis and most occurrences of supressor mutations result from competitive systems such as must be involved in the case just described, but such a possibility must be borne in mind.

* Representing work supported in part by a grant from the Rockefeller Foundation, and in part by a grant from the American Cancer Society upon the recommendation of the Committee on Growth of the National Research Council.

[1] Emerson, S., "Growth Responses of a Sulfonamide-Requiring Mutant Strain of *Neurospora*," *J. Bact.*, **54**, 195–207 (1947).

[2] Zalokar, M., "The *p*-Aminobenzoic Acid Requirement of the Sulfonamide-Requiring Mutant Strain of *Neurospora*," *Proc. Nat. Acad. Sci.*, **34**, 32–36 (1948).

[3] Beadle, G. W., and Coonradt, V., "Heterocaryosis in *Neurospora crassa*," *Genetics*, **29**, 291–308 (1944).

[4] Jones, D. F., "Heterosis Resulting from Degenerative Changes," *Ibid.*, **30**, 527–542 (1945).